DeepKey: An EEG and Gait Based Dual-Authentication System

XIANG ZHANG, University of New South Wales
LINA YAO, University of New South Wales
KAIXUAN CHEN, University of New South Wales
XIANZHI WANG, University of New South Wales
QUAN Z. SHENG, Macquarie University
TAO GU, RMIT University

Biometric authentication involves various technologies to identify individuals by exploiting their unique, measurable physiological and behavioral characteristics. However, biometric authentication systems (e.g., face recognition, iris, retina, voice, and fingerprint) are increasingly facing the risk of being tricked by biometric tools such as anti-surveillance masks, contact lenses, vocoders, or fingerprint films. In this regard, we design a multimodal biometric authentication system named DeepKey which uses both gait and Electroencephalography (EEG) signals to provide better protection against such risks. DeepKey consists of three key components: an Invalid ID Filter Model to block invalid subjects, a Gait Identification Model to recognize Gait IDs and an EEG Identification Model to recognize EEG IDs. In particular, the first two models employ a one-class SVM algorithm and a Recurrent Neural Network based deep learning model, respectively. The third model combines autoregressive coefficients, an RNN structure, and an SVM classifier. DeepKey is trained with a gait dataset of 160,000 samples and an EEG dataset of 108,000 samples. Experimental results show DeepKey outperforms a series of comparison methods and achieves an overall accuracy of 0.983 along with an overall false acceptance rate (FAR) of 0.0 and a false rejection rate (FRR) of 0.019.

CCS Concepts:
- Security and privacy → Biometrics; Biometrics;
- Computing methodologies → Machine learning algorithms;

General Terms: Gait, EEG, Biometric authentication System, Deep Learning

Additional Key Words and Phrases: Gait, EEG, biometric authentication, multimodal, deep learning

ACM Reference format:
Xiang Zhang, Lina Yao, Kaixuan Chen, Xianzhi Wang, Quan Z. Sheng, and Tao Gu. 2017. DeepKey: An EEG and Gait Based Dual-Authentication System. ACM J. Comput. Cult. Herit. 9, 4, Article 39 (March 2017), 20 pages.
DOI: 0000001.0000001

1 INTRODUCTION
Over the past decade, biometric authentication systems have gained more acceptance due to its reliability and adaptability. Existing biometric authentication systems generally include physiological and behavioral ones. The former is based on individuals’ unique intrinsic features (e.g., face [14], iris [29], retina [40], voice [15], and fingerprint [46]) and the latter is based on individuals’ behavior patterns such as gait analysis [5] and mobile phone usage pattern [52]. Within these systems, EEG signal-based cognitive biometrics and gait-based systems probably represent the most advantageous approaches.
The EEG signal-based system is an emerging approach in physiological biometrics in recent years. Such systems measure brain’s response to a number of stimuli in the form of EEG signals, which record the electromagnetic, invisible, and untouchable electrical neural oscillations. Lots of research efforts have been made on EEG-based biometric authentication. For instance, Chuang et al. [8] propose a single-channel EEG-based authentication system, which achieves an accuracy of 0.99. Sarineh Keshishzadeh et al. [23] employ a statistical model for analyzing EEG signals and achieves an accuracy of 0.974. Generally, EEG signals have the following inherent advantages:

- **Uniqueness.** 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 ingenious enough to protect against faked identities[8].
- **Reliability.** An EEG-based authentication system can prevent the subjects under abnormal situations (e.g., dramatically spiritual fluctuating, hysterical, drunk, or under threaten) since EEG signals are not 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, the gait-based authentication systems have been an active direction for years. Gait data are more generic and can be gathered easily from the popular inertial sensors. Gait data are also unique because they are determined by intrinsic factors (e.g., gender, age, height, limb length), temporal factors [6] (e.g., step length, step width, 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 long term and therefore difficult to be faked. Hoang et al. [20] propose a gait-based authentication biometric system to analyze the gait data gathered by mobile device, 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 gait patterns to detect invalid subjects (authenticated people). The proposed method achieves an EER (equal error rate) of 2.9%.

Although both the Gait-based and EEG-based authentication systems have promising characteristics, all those systems face the threat of being deceived as a result of the rapidly development of manufacturing industry and technologies. For example, people can easily trick a fingerprint-based authentication system by using a fake fingerprint film [1] or an expensive face recognition-based authentication systems by simply wearing a 200 dollars’ worth anti-surveillance mask [2]. Besides, various other challenges still remain: (i) the performance of the gait-based authentication system does not scaled well with the increase in the number of authorized subjects [42]; (ii) many biometric authentication systems are working on a single modality, the FAR of which is higher than 0.03. It is not precise enough for high degree confidential places such as military bases, the treasuries of banks and political offices which require an authentication system with high precision since any tiny misjudge will provoke great economic or political catastrophes; (iii) the authentication system may result in wrong decisions if disturbed by the environmental factors.

In this paper, we propose a two-factor biometric authentication system, DeepKey, which enables the dual-authentication to leverage the advantages of both the gait-based and the EEG-based systems. Compared with either gait-based or EEG-based authentication system, the two-factor authentication system offers more reliable and precise identification. Table 1 summarizes the overall comparison of DeepKey with some representative works on 7 key aspects. DeepKey consists of three main components: the Invalid ID Filter Model to eliminate invalid or outlier targets; the Gait Identification Model to identify targets’ gait IDs; and the EEG Identification Model
Table 1. Comparison of various biometrics. EEG and Gait have considerable fake-resistance which is the most significant character of authentication systems. ↑ denotes the higher the better while ↓ denotes the lower the better.

| Biometrics  |Fake-resistance |Universality |Uniqueness |Stability |Accessibility |Performance |Computational cost |
|-----------|----------------|-------------|-----------|----------|-------------|------------|------------------|
| Face/Vedio | Medium         | Medium      | Low       | Low      | High        | Low        | High             |
| Fingerprint/Palmprint Iris |Low |High |High |High |Medium |High |Medium |
| Uni-modal Retina |High |High |Medium |Medium |Low |High |High |
| Voice |Low |Medium |Low |Low |Medium |Low |Low |
| Gait |High |Medium |High |Medium |Medium |High |Low |
| EEG |High |Low |High |Low |Low |High |Low |
| Multi-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 |High |Medium |
| ECG+fingerprint |High |Medium |Medium |High |Low |High |Medium |
| DeepKey |High |Low |High |High |Low |High |Low |

to identify targets’ EEG IDs. A target is granted access only after she/he passes two separate authentications. Our main contributions of this paper are highlighted as follows:

- We present DeepKey, a dual-authentication system that exploits both gait and EEG biological traits. To the best of our knowledge, DeepKey is the first two-factor authentication system which combines EEG and gait signals for person authentication. To further enhance the performance, DeepKey includes an Invalid ID Filter Model that achieves a FAR of 0 to eliminate impostors with 100% security guaranteed.
- We design a robust recurrent neural network that simultaneously detection and classification of multi-modal sensor data, to decode the large diversity in how people perform gaits and how people perform brain activities.
- Together a cropped training strategy, Orthogonal Array Experiment Method, we validate and evaluate DeepKey on real-datasets. Our results show that DeepKey significantly outperforms a series of baseline models and the-state-of-the-art methods, achieving a FAR of 0 and a FRR of 0.019.

Note that all the necessary reusable codes and datasets in this paper have been open-sourced for reproduction, please refer to this link. 3

The remainder of this paper is organized as follows. Section 2 introduces the EEG-based, gait-based, and multimodal biometric systems briefly. Section 3 details the methodology and three key models (Invalid ID Filter Model, Gait Identification Model, and EEG Identification Model) of the DeepKey authentication system. Section 4 evaluates the proposed approach on the public Gait and EEG dataset and provides analysis of the experimental results. Finally, Section 5 summarizes this paper and points out our future work.

2 RELATED WORK

2.1 Biometric authentication technologies

Since biometric features cannot be stolen or duplicated easily, biometric authentication [4] is becoming increasingly a commonplace. Currently, the most mature biometric authentication technology is fingerprint-based authentication which has been demonstrated to high matching accuracy and been used for decades [32]. Iris recognition is another popular approach for biometric authentication owing to its unique and stable pattern [39]. In 1993, 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 [39] has emerged offering solutions to ameliorate iris authentication problems. For example, Pillai et al. [39] introduce

3https://github.com/xiangzhang1015/DeepKey
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 [12, 18, 53] are the most common used and accepted by the public for its unique features and non-intrusiveness. Since face recognition systems require tackling different challenges including expression, image quality, illumination, and disguise to achieve high accuracy, infrared (IR) [18] and 3D [12] systems have attracted much attention. According to [53], multimodal recognition combining traditional visual textual features and IR or 3D systems can achieve higher accuracy than single modal systems.

2.2 Gait based authentication

As the most basic activity in our daily lives, walking is an advanced research hotspot for activity recognition [21, 36, 51]. Differing from previous work, our work focuses on human gait, a spatiotemporal biometric that measures a person’s manner on walking. On the one hand, gait can be collected remotely without human interaction when compared with other aforementioned biometric features [47]. On the other hand, it is challenging to eliminate the influence of exterior factors including clothing, walking surface, shoes, carrying stuff, and other environmental factors. Existing gait recognition approaches sit in two categories. One is model-based approach [38], which models gait information with a 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) [33] has become one of the most popular appearance-based methods in recent years. Based on GEIs, Guan et al. [16] employ a Random Subspace Method (RSM) [19] and model the influence of exterior factors as the unknown partial feature corruption to decrease the influence. With the prevalence of WiFi devices, WiFiU [48], a method utilizing two WiFi devices to extract discriminating gait features is proposed. To address the challenges of distinguishing different body parts and representing walking patterns, the system generates spectrograms and conduct enhancing process and autocorrelation on the contour profile. Besides, the cross-view variance is also a concern of gait identification [27, 28]. For example, Wu et al. [49] consider not only the cross-view variance but also deep convolutional neural networks (CNNs) for robust gait identification.

2.3 EEG based authentication

Since EEG can be gathered in a safe and non-intrusive way, researchers have paid great attention to exploring this kind of brain signals. For person authentication, EEG is promising for being confidential and fake-resistant but on the other hand, complex and hard to be analyzed. Marcel and Millán [35] use Gaussian Mixture Models and train client models with MaxIMUsm A Posteriori (MAP). Ashby et al. [2] extract five sets of features from EEG electrodes and inter-hemispheric data, combine them together, and process the final features with support vector machine (SVM). The study shows that EEG authentication is also feasible with less-expensive devices. Altahat et al. [1] select Power Spectral Density (PSD) as the feature instead of the widely used autoregressive (AR) models to get higher accuracy. They also conduct channel selection to determine contributing channels among all 64 channels. Thomas and Vinod [44] take advantage of individual alpha frequency (IAF) and delta band signals to compose specific feature vector. They also prefer PSD features but only perform the extraction merely on gamma band.

2.4 Multimodal biometric authentication

Since traditional unimodal authentication suffers from the negative influence of loud noise, low universality, and intra-class variation, it can no longer meet the higher accuracy requirement by a wide range of applications. To address this concern, multimodal biometric authentication which combines and uses biometric traits in different ways becoming popular. Taking the commonness into consideration, most works choose two biometrics from face, iris, and fingerprints and make the fusion [24, 26, 43]. In [11], an innovative combination between gait and
electrocardiogram (ECG) is shown to be effective. Manjunathswamy et al. [34] combine ECG and fingerprint at the score level. To the best of our knowledge, the approach proposed in this paper is the first one to combine EEG and gait for person authentication. Taking advantages of both EEG and gait signals, the combination is expected to improve the reliability of any authentication systems.

2.5 Summary
As a summary, traditional authentication techniques only consider common and easily accessible biometrics devices. Nevertheless, with advanced scientific technologies, these biometrics systems are exposed to the danger of being faked. Relying upon the fake-resistance power of both approaches, an innovative combination of EEG and gait guarantees the safety and correctness of authentication systems. Furthermore, most of the works do not mention their solutions to dealing with invalid subjects (e.g., an imposter). On the contrary, our work presents a complete Invalid ID Filter Model and achieves a high Invalid ID Filter performance.

3 THE PROPOSED APPROACH
In this section, we first give an overview of the DeepKey system and then present the technical details for each component, namely Invalid ID Filter, Gait-based Identification, EEG-based Identification and Decision Making.

3.1 System Overview
The DeepKey system is supposed to be deployed in front of the access to the confidential location (e.g., bank vouchers, military bases and government confidential residences). It includes an aisle for gait data collection and an EEG helmet for EEG data collection (as shown in the left part of Figure 1). The procedure of DeepKey can be described as follows. The person goes through the aisle for gait data collection and puts on the EEG helmet for EEG data collection at the end of the aisle. After that, the final decision (Approval or Rejection) can be made according to both the gait and the EEG identification results.

All data including the ground truth (the ID label of the target subject), gait component, and EEG data component are input into the DeepKey system for future analysis:

\[ \text{Inputdata} = \{ID|[Gaitdata : EEGdata]\} \]

As shown in Figure 1, the analysis contains four steps:

1. Firstly, the Invalid ID Filter Model judges based on the gait data if the subject is an impostor or a genuine person. If the answer is yes (impostor), the Decision Making step will deny the request. The algorithm for the Invalid Detection Model is a one-class SVM (Section 3.2).

2. If the individual is determined as genuine, the Gait Identification Model will identify the individual’s authorized ID based on the gait data. This model is pre-trained off-line with the Recurrent Neural Networks (RNN) based model (Section 3.3). The output is the ID number associated with the person’s detailed personal information.

3. The third step is EEG Identification. Similar to the prior step, this step identifies the subject’s EEG ID based on the EEG input data. The pre-trained EEG ID Model consists of three components: Autoregressive coefficient, RNN and SVM (Section 3.4).

4. The final step is to check the consistency of the gait ID and the EEG ID. If they are identical (e.g., both ID numbers are 2), the decision-making section will grant an approval, otherwise deny the subject and take corresponding security measures.

One key component of DeepKey is Invalid ID Filter Model as both the Gait Identification Model and the EEG Identification Model do not contain a threshold to filter invalid outlier subjects. After the person goes through the aisle for gait data collection and puts on the EEG helmet for EEG data collection at the end of the aisle, the decision...
(Approval or Rejection) can be made according to both the gait and the EEG identification results. Therefore, the filter section takes the responsibility for avoiding misidentifying invalid subjects as genuine. Another one is Recurrent Neural Networks (RNN), the core algorithm of DeepKey. With deep architectures, DeepKey can avoid the time-consuming pre-processing and feature extraction in data processing. Moreover, RNN plays two important roles in Algorithm 1 and Algorithm 2. RNN mainly takes classification function in Algorithm 1 and works as a feature extraction machine and sends refined features to SVM for classification in Algorithm 2. All the important parameters mentioned in this paper are listed in Table 2.

3.2 Invalid ID Filter Model

The subjects in an authentication system are categorized into two classes: authorized and unauthorized. We apply one-class SVM to sort out the unauthorized subjects.

Given a set of authorized subjects $S = \{S_i, i = 1, 2, \cdots, L^o\}, S_i \in \mathbb{R}^{n_s}$, where $L^o$ denotes the number of authorized subjects and $n_s$ denotes the number of features in input data. The input data consists of gait data $G = \{G_i, i = 1, 2, \cdots, L^o\}, G_i \in \mathbb{R}^{n_g}$ and EEG data $E = \{E_i, i = 1, 2, \cdots, L^e\}, E_i \in \mathbb{R}^{n_e}$. $n_g$ and $n_e$ denote the number of features in gait data and EEG data, respectively, and

$$n_s = n_g + n_e$$

In the Invalid ID Filter Model, only the gait data $G_i$ is utilized. The $\mathbb{R}^{n_g}$ is called the input space. A nonlinear function $\Phi(G_i)$ maps vector $G_i$ to a higher dimensional feature space $F$, which provides a hyper-plane

$$f(G) = \left\{ \sum_{i=1}^{L^o} \alpha_i K(G, G_i) - \rho \right\}$$

as the boundary of the mapped vector $\{\Phi(G_i), i = 1, 2, \cdots, L^o\}$ to enclose as many samples as possible into a hyper sphere area with a minimum volume. $\alpha_i$ denotes the Lagrange multiplier obtained by optimizing

$$\min_{\alpha} \left\{ \frac{1}{2} \alpha_i \alpha_j K(G_i, G_j) \right\}, s.t. 0 \leq \alpha_i \leq \frac{1}{v m}, \sum_{i=1}^{L^o} \alpha_i = 1$$

$\rho$ denotes the distance between the feature space and the input space. $v$ denotes the threshold, which determines the sensitivity of the model to invalid samples. $m$ is the number of training data. The kernel function is an

![Fig. 1. Work-flow of the Proposed Approach. The picture comes from Mission Impossible.](image-url)
DeepKey: An EEG and Gait Based Dual-Authentication System

Table 2. Notation

| Parameters | Explanation |
|------------|-------------|
| \(S_i\)    | the \(i\)-th input data |
| \(n_s\)    | 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 |
| \(L^o\)    | the number of input data samples |
| \(\alpha_i\) | the Lagrange multiplier |
| \(\nu\)    | the threshold in one-class SVM |
| \(m\)      | the number of training data |
| \(\gamma\) | the parameter in RBF kernel |
| \(m_{in}\) | the number of rows of gait training input data |
| \(bs\)     | batch size |
| \(lr\)     | learning rate |
| \(n_k\)    | the number of nodes in the \(k\)-th layer of gait Identification Model |
| \(X_k\)    | the data in \(k\)-th layer in deep learning model |
| \(b_k\)    | the biases of \(k\)-th layer in deep learning model |
| \(Y_j\)    | the prediction label of \(j\)-th sample |
| \(p\)      | the rows of input EEG data |
| \(r\)      | the number of fragments in Input EEG data |
| \(q\)      | the number of EEG samples in each fragment |
| \(x[h]\)   | the \(h\)-th element in one column in the EEG fragment |
| \(n_e\)    | the number of EEG raw data columns |
| \(\theta\) | the order of autoregressive |
| \(\alpha_h\) | the corresponding coefficients of \(x[h]\) |

The kernel function is defined as a RBF kernel,

\[ K(G, G_i) = \Phi(G)^T \Phi(G_i) \]

where \(d(G, G_i)\) is the distance between the learned common pattern \(G\) and a specific sample \(G_i\). \(\gamma\) denotes the parameter of RBF kernel. The test samples will be accepted as normal samples if \(f(G) > 0\); otherwise, they will be regarded as invalid subjects. The testing data are divided into a number of fragments where each fragment includes a number of samples (rows). The final judgment result is the number of the judgment on all the samples. For example, given a fragment of 200 samples, if 150 of them are classified to be normal and 50 to be invalid, then the final result of this fragment is normal.

3.3 Gait Identification Model

RNN is a supervised deep learning algorithm that performs well in classification and regression models. It gives each neuron and neuron connection a modifiable real-valued weight and explores the mapping relationship between the input features and the expected output results. By exploiting the time series relationship between
samples and adjusting the weights of all the neurons of all layers, the RNN structure can imitate any mapping relationship. The RNN structure contains three components: the input layer, the hidden layer and the output layer.

The input layer receives the input features, with each node corresponding to one feature. The hidden layer can be one or more layers. Specially, a model is considered a deep learning model when its hidden layer is more than one layer, where more nodes and layers can potentially enhance the performance of the mapping between the input layer and the output layer. The output layer produces the results of the algorithm with each node in this layer corresponding to one output feature.

In our system, the Gait Identification Model employs a 7-layer RNN model including an input layer, five hidden layers, two of which are composed of LSTM (Long Short Term Memory), and an output layer.

Suppose \( G \) is the input data of the Gait Identification Model and it has the shape of \([n_{in}, n_g]\), where \( n_{in} \) denotes the rows of gait data. The input data are divided into several mini-batches for less memory occupation since every iteration has fewer samples and faster training since the parameters of the model update after every iteration. \( bs \) denotes the number of samples in each mini-batches.

Parameters of the model include \( lr \) which denotes the learning rate in the learning model and \( n_k, k \in \{1, 2, \cdots , 7\} \) which denotes the number of the nodes in the \( k \)-th layer. The output of the deep learning model is one hot ID label of the subjects which means that the label of a sample is represented by a list of binary units, \( L \) units.

The data in the \( k \)-th layer are denoted by \( X_k \in R^{[n_{in} \times n_k]}, k \in \{1, 2, \cdots , 7\} \). For instance, \( X_1 \) denotes the data in layer 1 (input layer). The weights between layer \( k \) and layer \( k + 1 \) are denoted by \( W_{k(k+1)} \in R^{[n_k \times n_{k+1}]}, k \in \{1, 2, \cdots , 6\} \), \( b_k \in R^{n_k} \) denotes the biases of \( k \)-th layer. Each input sample can be 2-D when \( n_s > 1 \) or 1-D when \( n_s = 1 \).

Suppose the input data are 3-dimensional tuples in the form of \([bs, n_s, n_{in}]\), which consist of \( bs \) samples in the form of \([n_s, n_{in}]\), they are reshaped to \([bs \times n_s, n_{in}]\).

\[
X_k = X_{k-1} \ast W_{k(k+1)} + b_k, k \in \{1, 2, \cdots , 6\} 
\]

Note that both the fourth and the fifth layers are LSTM layers and the sizes of \( X_k \), \( W_{k(k+1)} \) and \( b_k \) must match.

To increase the non-linearity of this model, we use a sigmoid function as the activation function. Through pre-experiments, we find the model performs the best when the sigmoid function works on the second layer. So \( X_2 \) becomes the following:

\[
X_2 = sigmoid(X_1 \ast W_{12} + b_1) 
\]

To make the difference between categories more apparent, a softmax function is employed on the results \( X_7 \) described by the following formula:

\[
X'_{ij} = \frac{X_{ij}}{\sum_{i=1}^{n_l} X_{ij}}, \quad i \in \{1, 2, \cdots , n_l\} 
\]

where \( X'_j \in R^{[bs \times n_l]}, X'_{kj} \) is the \( i \)-th value in the \( j \)-th sample’s outputs (the classification result of the \( j \)-th sample), \( n_l \) is the number of the elements in a sample label, \( n_l = n_c + 1 \). Besides, the nodes of the output layer and the elements in the single label should be of the same number, i.e., \( n_l = n_K \). We choose the objective as a Log loss function with a modified L2 norm loss \(^4\) to avoid over-fitting.

\[
\text{cost} = -\frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{n_l} (y_{ij} \log(X'_{kij}) + (1 - y_{ij}) \log(1 - X'_{kij})) + L2 
\]

\(^4\)https://github.com/tensorflow/tensorflow/blob/master/tensorflow/g3doc/api_docs/python/functions_and_classes/shard4/tf.nn.l2_loss.md
where $N$ is the number of samples in this batch ($N = bs$), $y \in \mathbb{R}^{(bs \times n)}$, $y_{ij}$ is the $i$-th element in the $j$-th sample’s true label. The modified $L2$ loss is calculated as:

$$L2 = \lambda \sum_{h=1}^{nh1} \frac{\sum_{h2=1}^{nh2} v_{h2}^2}{2}$$

where $\lambda$ is the coefficient, $v$ is trainable variables in the network, $n_{h1}$ is the number of variable tensors in the network, and $n_{h2}$ is the number of values in each specific tensor. The Adam optimizer is employed to optimize the cost. This optimizer uses moving averages of the parameters and enables to choose a bigger effective step and convergence to the step size even without fine tuning.

The forecast results ($X^*_t$) of the model is a list of values of length $n_t$. For example, suppose one single sample’s classification result is $[0.067, 0.134, 0.067, 0.402, 0.052, 0.134]$, and we use an argmax function to find the position of the maxIMUSm and get the largest value 0.402 at the position 3 (the position is counted from zero). The prediction result of this sample is also 3. If the true label of this sample is 3, the number of true labels is added by 1; otherwise, by 0. At the end, the accuracy of this batch samples is the average of all counts:

$$acc = \frac{\sum_{j=1}^{bs} a_j}{bs}$$

$$a_j = \begin{cases} 
1 & Y_j = label_j \\
0 & other 
\end{cases}$$

where $label_j$ is the true label of the $j$-th sample.

The common structure of $L$-layer LSTM will be introduced ($L=2$ in this paper). Each layer of the Multilayer LSTM is connected by several single LSTM cells. The output of the cell is not only related to the input of this cell but also related to the previous cell’s output. The multilayer LSTM layers are connected by $L$ layers of LSTM cells. Every layer should have the same number of basic cells. For the cells in the same order in different layers, the input of the later layer’s cell equals to the output of the previous layer’s cell, represented by $I^t = O^{t-1}_{t-1}$. Every LSTM cell (e.g., Cell $t$ in Layer $L$) has three inputs (two from the same layer (e.g., $O^{t}_{t-1}$, $c^{t}_{t-1}$) and one from previous layer (e.g., $O^{t-1}_{t-1}$)) and two outputs ($c^{t}_{t}$ are sent to the $(t+1)$-cell in Layer $L$, while $O^{t}_{t}$ is sent to both the $(t+1)$-cell in Layer $L$ and the $t$-cell in Layer $L+1$).

Suppose $O^{t}_{t}$ is the output of the $t$-th LSTM cell in Layer $L$, $c^{t}_{t}$ denotes the state (memory) of the $t$-th LSTM cell in $L$ layer, $I^{t}_{t}$ denotes the Input of the $t$-th LSTM cell in $L$ layer, $T$ denotes an operation ($Wx + b$) for some weights $W$ and biases $b$, and $\circ$ denotes element-wise multiplication, since $L$ is the number of LSTM cell layers. A LSTM cell has three inputs and two outputs as follows:

$$LSTMcell : O^{L}_{t}, c^{L}_{t} \leftarrow O^{t-1}_{t}, O^{t-1}_{t}, c^{t-1}_{t}$$

The specific equations are described as the following:

$$i = sigmoid(T(O^{t-1}_{t}, O^{t-1}_{t}))$$

$$f = sigmoid(T(O^{t-1}_{t}, O^{t-1}_{t}))$$

$$o = sigmoid(T(O^{t-1}_{t}, O^{t-1}_{t}))$$

$$m = tanh(T(O^{t-1}_{t}, O^{t-1}_{t}))$$

$$c^{t}_{t} = f \circ c^{t-1}_{t} + i \circ g$$

$$O^{t}_{t} = o \circ tanh(c^{t}_{t})$$

\[\text{References}\]

1. Supposed

2. https://www.tensorflow.org/versions/r0.10/api_docs/python/rnn_cell/rnn_cells_for_use_with_tensorflow_s_core_rnn_methods

3. BasicLSTMCell

ACM J. Comput. Cult. Herit., Vol. 9, No. 4, Article 39. Publication date: March 2017.
ALGORITHM 1: Gait Identification Model

Input: Gait data $G$ and corresponding parameters $lr, bs, \lambda, n_k, n_{iter}$
Output: Gait ID number, $Y_j$
1: while iteration < $n_{iter}$ do
2:   for $k \in \{1, 2, \cdots, 7\}$ do
3:       $X_k \leftarrow f(X_1, W_{(k+1)j}, b_k)$
4:       $Y_j \leftarrow \text{argmax}(\text{softmax}(X_{1j}))$
5:   end for
6: end while
7: Gait ID = $Y_j$
8: return Gait ID

where $i, f, o$ and $m$ represent the input gate, forget gate, output gate and input modulation gate \(^7\), respectively. Each gait sample is a single row vector in the Gait Identification Model. We set state_is_tuple as True, and activation as None.

3.4 EEG Identification Model

Compared to gait data, the EEG data contain more noise which are more challenging to handle. Given the complexity of EEG signals, the data pre-processing is necessary. The EEG Identification Model is constructed by three components: AR for pre-processing, RNN for feature extracting, and SVM for classification. EEG signals are time series signals, which mean that the current samples’ EEG value is related to the values of previous samples and the relationship of EEG signal differs from one subject to another. Auto-regressive Coefficients (AR) is one of the most widely used pre-processing methods on EEG data. The classical classifier SVM is employed with RNN. The RNN in EEG Identification Model has the same structure as the RNN model in the Gait Identification Model (details in Section 3.3). However, the function of RNN structure in the EEG Identification Model is to redefine the feature and extract the useful and dominate features so that the SVM classifier can clearly distinguish them.

To distinguish the parameters in two RNN structures, subscript $\epsilon$ is added to the EEG Identification Model’s RNN parameters. For example, in this RNN model, suppose $lr_\epsilon$ denotes learning rate, $bs_\epsilon$ denotes batch size, $\lambda$

\(^7\)More details about the LSTM structure can be found in http://colah.github.io/posts/2015-08-Understanding-LSTMs/
ALGORITHM 2: EEG Identification Model

Input: EEG raw data $E$ and corresponding parameters $p$, $q$, $\theta$, $lr$, $bs$, $\lambda$, $n_{k,e}$, $n_{iter,e}$

Output: EEG ID number

1. for $q$ do
2. Autoregressive Coefficients, $C \leftarrow E$
3. end for
4. while iteration $< n_{iter,e}$ do
5. for $k \in \{1, 2, \cdots, 7\}$ do
6. $X_{k,e} \leftarrow f(X_{1,e}, W_{k,e(k,e+1)} e, b_{k,e})$
7. $Y_{j,e} \leftarrow \text{argmax}(\text{softmax}(X_{jij,e}))$
8. end for
9. end while
10. for $Y_{j,e}$ do
11. Linear SVM, EEG ID number $\leftarrow Y_{j,e}$
12. end for
13. return EEG ID

denotes the coefficient of $L2$ norm, $n_{k,e}$ denotes the number of neurons in the $k_{e}$-th layer and $n_{iter,e}$ denotes the iterations of learning progress. A linear SVM classifier is employed to classify the subject’s ID number of input EEG data.

Up to this point, the Decision Making module makes the final decision based on both the Gait ID and the EEG ID results. The decision of Approval is granted only when the input data successfully pass the Invalid ID Filter and the Gait Identification and the EEG Identification models identify the subject as the same person. Otherwise, Rejection is declared.

4 EXPERIMENTS AND RESULTS

4.1 Experimental Settings

The proposed DeepKey system is trained and evaluated using two public datasets: a Gait dataset and an EEG dataset.

**Physical Activity Monitoring Dataset.** The Gait data is extracted from a PAMAP2 physical activity monitoring dataset 8 of the UCI Machine Learning Repository. 18 different physical activities performed by 9 subjects who wear 3 inertial measurement units (IMUs) and a heart rate monitor. The data are collected with sampling rates of 100Hz from devices deployed over the wrist on the dominant arm, chest and the dominant side’s ankle. Each IMU contains 5 sensors and totally 17 channels: a temperature sensor with 1 channel, two 3D-acceleration sensors with 3 channels each, a 3D-gyroscope sensor with 3 channels, 3D-magnetometer with 3 channels and an orientation sensor with 4 channels. Since most of heart rate data are NaN, we select 3 IMUs with totally 51 channels and 8 subjects, each with 20,000 walking gesture data to evaluate our DeepKey method.

**EEG Motor Movement/Imagery DataBase.** The EEG data used in this paper come from PhysioNet eegmmidb (EEG motor movement/imagery database) dataset 9, which is collected with the BCI2000 (Brain Computer Interface) instrumentation system 10 [41]. The eegmmidb dataset contains around 26.4 million samples collected from 109 subjects and the BCI 2000 system has 64 channels and the sampling rate is 160 Hz. According to the tasks,
different annotations are labeled and can be downloaded from PhysioBank ATM. 11 The actions in different tasks are as follows:

- **Task 1**: A target appears on either the left or the right side of the screen. The subject imagines to open and close the corresponding fist as the target appears and disappears.

- **Task 2**: A target appears on either the top or the bottom of the screen. The subject imagines open and close either both fists (if the target is on top) or both feet (if the target is on the bottom) until the target disappears.

In these tasks, the condition is labeled as the rest representing the relax state of subjects. 108,000 samples collected in a calm state of 8 subjects, that is, 13,500 for each, are selected for our experiments. Every sample is one list including 64 elements corresponding to 64 channels.

### 4.2 System Evaluation

We extensively evaluate the DeepKey system, along with three key components, Invalid ID Filter, Gait ID Identification, and EEG ID Identification, respectively.

Generally, the reliability and performance of biometric systems are measured by two evaluation standards: *False Acceptance Rate* (FAR) which means the rate of the system accepting impostors as genuine, and *False Rejection Rate* (FRR) which means the rate of the system rejecting genuine subjects as impostors. Because of the threshold in the invalid ID filter machine, both FAR and FRR are continuous lines along with the change of threshold. FAR and FRR are calculated as:

\[
FAR = \frac{\text{impostor numbers exceeding threshold}}{\text{all impostor numbers}}
\]

\[
FRR = \frac{\text{genuine numbers exceeding threshold}}{\text{all genuine numbers}}
\]

For instance, assuming that we have 100 impostor samples, 90 of which are classified to be impostors while 10 of which are falsely classified to be genuine. Similarly, we have 200 genuine samples, 150 of which are classified correctly while 50 of which are mistakenly classified to be impostors. In this case, \(FAR = 50/200 = 0.25\) and \(FRR = 10/100 = 0.1\).

**The Results of Invalid ID Filter.** In the Invalid ID Filter experiment, we select 6 subjects randomly to be genuine subjects and as the training set while the other two to be impostors and as the test set. In our system, the Invalid ID Filter Model of DeepKey achieves an FAR of 0, denoted by \(P_{FAR}\), and a FRR of 0.0036, denoted by \(P_{FRR}\). It should be noted that for a confidential authentication system, FAR is always regarded more important than FRR and the zero FAR means all the invalid subjects can be blocked out by DeepKey.

**The Results of DeepKey.** For the whole system, the subjects passing the Invalid ID Filter are regarded as genuine only if their recognized IDs are consistent, that is, \(Gait\ ID = EEG\ ID\). It can be inferred easily that the FAR of DeepKey is 0 as well. However, for FRR, the situation is more complex. The false reject decision depends on one or more of these three components: the false rejection of Invalid ID Filter, the incorrect classification of Gait Identification, and the incorrect classification of EEG Identification. For \(P_{FRR}\) denotes the Invalid ID Filter’s FRR and \(P_{FAR} = 0\); \(P_{FRR}\) denotes the rate of the Invalid ID Filter working correctly, and \(P_{GF}, P_{GT}, P_{EF}\) and \(P_{ET}\) denote the rates of false classification of Gait ID, true classification of Gait ID, false classification of EEG ID and true classification of EEG ID, respectively. Therefore, the FRR of DeepKey is calculated by

\[
FRR = P_{FRR} + P_{FRR} \times (P_{GF} \times P_{ET} + P_{GT} \times P_{EF} + P_{GF} \times P_{EF})
\]

Another evaluation standard, EER (Equal Error Rate), refers to the point at which the threshold leads to FAR = FRR, which therefore means the biometric system gets the best trade-off between rejection and acceptance.

11https://www.physionet.org/cgi-bin/atm/ATM
Table 3. Evaluation of DeepKey system and its components

| Item               | Invalid ID Filter | Gait Identification | EEG Identification | DeepKey          |
|--------------------|-------------------|---------------------|--------------------|------------------|
| $P_{FAR}$          | 0.0036            | 0.9964              | 0.999              | 0.00159          |
| $P_{FRR}$          | 0.999             | 0.78                | 0.85               | 0.77             |
| $P_{GT}$ (Acc)     | 0.9841            | 0.984               | 0.984              | 0.983            |
| $P_{GF}$ (Acc)     | 0.0159            | 0.015               | 0.014              | 0.019            |
| Rate               | 0                 | 0.996               | 0.999              | 0.019            |

Table 4. Comparison among classification methods. Different methods are variously suitable for different categories of datasets. All the methods use the same EEG and gait datasets.

| Gait ID Method     | RNN               | DWT(a)+AE+XGB       | DWT(aa)+AE+XGB     | XGB              | DWT(aa)+AE+RNN   |
|--------------------|-------------------|---------------------|--------------------|------------------|------------------|
| Accuracy           | 0.999             | 0.72                | 0.78               | 0.58             | 0.4              |

| EEG ID Method      | AR+RNN+SVM        | AR (0.5)+RNN+SVM    | RNN                | AE+XGB           | RNN+AE+XGB       |
|--------------------|-------------------|---------------------|--------------------|------------------|------------------|
| Accuracy           | 0.984             | 0.84                | 0.85               | 0.77             | 0.75             |

The lower the EER is, the better the system achieves. In our system, for FAR remains to be 0, only FAR and a selected FRR are used for evaluation, while existing research use EER together with FRR and FAR as evaluation measurements.

All the variables can be found in Table 3. In terms of the performance of the DeepKey system, it obtains the FAR of 0 and the FRR of 0.019, which ensures that the DeepKey system will never grant an incorrect person with authority and the probability of the authorized subjects being blocked out is approximately 0.019. Since DeepKey includes both Gait and EEG Identification components, which achieves the accuracy of 0.999 and 0.984, and the overall accuracy of the fusion system is $0.999 \times 0.984 \approx 0.983$.

4.3 Comparison

In this section, we compare the performance of DeepKey with several other classification methods. For Gait classification, the proposed Gait Identification Model includes a 7-layer RNN structure while for EEG classification, the EEG Identification Model mainly contains Auto-regressive coefficient extraction, feature extraction with RNN and classification with SVM. The performance of related approaches are listed in Table 4.

In Table 4, XGB means the XGBoosting method [7]. DWT\(^{12}\) (Discrete Wavelet Transform) is the wavelet transformation with the wavelets discretely sampled. DWT is a data processing method which enables to capture both frequency and time-domain information. AE\(^{13}\) (Autoencoder) is an unsupervised neural network which is used in feature extraction and generally for dimensionality reduction. DWT(a) means one layer wavelet decomposition while DWT(aa) denotes two layers. Auto denotes auto-regressive coefficients. AR means Auto-regressive coefficients and AR (0.5) means there is a 50% overlap in the segmentation of EEG samples. All the methods working on Gait data are evaluated by the same dataset, so as EEG data.

Table 5 shows the comparison between the state-of-the-art authentication systems. Electrocardiography (ECG) describes the electrical activity of hearts over a period of time. The results show that the Gait Identification Model (accuracy of 0.999) and EEG Identification (accuracy of 0.984) outperform their counterparts, respectively. The proposed DeepKey authentication system gains the lowest FAR (0) and the acceptable FRR (0.019). Although some literature achieves slightly lower FRR than DeepKey, their FAR is much higher. As mentioned before, FAR has higher significance than FRR for a confidential chamber. Overall, DeepKey has better performance than the state-of-the-art authentication systems.

\(^{12}\)https://en.wikipedia.org/wiki/Discrete_wavelet_transform

\(^{13}\)https://en.wikipedia.org/wiki/Autoencoder

ACM J. Comput. Cult. Herit., Vol. 9, No. 4, Article 39. Publication date: March 2017.
Table 5. Comparison between DeepKey and other biometrics authentication systems. NN means Nearest-Neighbor; ANN means artificial neural network; AVTM-PdVs means extended view transformation models with a part-dependent view selection scheme; PSD means power spectral density; ECOC means error-correcting output code multi-class model; IAF means individual alpha frequency; CSP means Common Spatial Patterns; LDA means Linear discriminant analysis; AR means Autoregressive coefficients; ZM means Zernike Moment; RBF means Radial Basis Function; DoG means Difference of Gaussians; and LBP means Local Binary Patterns.

| Reference | Biometric traits | Method | Accuracy | FAR | FRR | EER |
|-----------|------------------|--------|----------|-----|-----|-----|
| Uni-modal | Gait             | semisupervised anomaly detection+NN | 97.4     | 2.9 |
| [9]       |                  | AVTM-PdVs | 77.72    | 2.91|
| [25]      |                  | Two SVMs | 1        |     |
| [45]      | EEG              | PSD + cross-correlation values | 90.03    | 0.22|
| [17]      | EEG              | Customized Threshold | 94.44    |     |
| [44]      | EEG              | IAF + delta band EEG + Cross-correlation values and mahalanobis distance measures | 90.21    |     |
| [22]      | EEG              | CSP + LDA | 96.97    |     |
| [23]      | EEG              | AR + SVM | 97.43    | 2.57|

| Multi-modal | Fingerprint | ZM + RBF Neural Network | 92.892 | 7.108 | 7.151 |
|             | Face         |                         | 73.2   | 11.52 | 13.47 |
| Fusion      |             |                         | 4.95   | 1.12  |       |
| Iris        | Gabor 2D wavelets + Gabor 2D wavelets + Hamming distance | 13.88  | 5.47  |       |
| Pupil       | Fusion       |                         | 2.44   | 2.44  | 2.44  |
| ECG         | wavelet decomposition | 2.38 | 9.52  |       |
| [50]        | Fingerprint  | Histogram manipulation Image Enhancement | 7.77   | 5.55  |       |
| Fusion      | Score Fusion |                      | 2.5    | 0     |       |
| [34]        | ECG           | linear time interpolation + cross correlation | 4.2    |       |       |
| Gait        | Fusion       | Score Fusion | 1.26   | 1.26  | 1.26  |
| Ours        | EEG          | AR+RNN+SVM | 99.9   |       |       |
| Fusion      |             |                         | 98.4   | 0     | 1.9   |

4.4 Invalid ID Filter Results

In this section, gait data, EEG data, and the combination of gait and EEG data from the PAMAP2 and eegmmidb dataset are separately used to train a one-class SVM model. These filters are called the gait-based filter, the EEG-based filter, and the combined filter, respectively. Still, 6 subjects are regarded as authorized subjects and the other 2 subjects are regarded as invalid subjects.

The gait based filter is established with 120,000 training gait samples and 40,000 test gait samples. No sample label is required since one-class SVM classifier is an unsupervised algorithm. To enhance the accuracy and robustness of the classifier, gait samples are separated into different gait segments, each segment with 200 continuous samples. All the 120,000 training gait samples are input to the one-class SVM classifier in training stage and 1,000 segments are randomly selected to test the performance in the test stage. The kernel in classifier is set as rbf, gamma = 0.1 and all of other parameters are set as default values but nu. Specially, nu (an upper bound on the fraction of training errors and a lower bound of the fraction of support vectors) is the most significant parameter in this classifier because it represents the classification threshold. Different nu value (from 0.01 to 1) is
DeepKey: An EEG and Gait Based Dual-Authentication System

The relationship between FRR, the training time and the size of segment

Fig. 2. The relationship between FRR, the training time and the size of segment

Fig. 3. FRR of Invalid ID Filter with error bars

Fig. 4. Accuracy of Gait Identification with error bars

Fig. 5. Accuracy of EEG Identification with error bars

Table 6. Invalid ID Filter Model performance while \( n_u = 0.15 \). True means authorized while False means unauthorized.

| Predict label | FAR   | FRR   |
|---------------|-------|-------|
| True          | 0.9964| 0.0036|
| False         | 0     | 1     |

tried and the performance shows that the FRR achieves the lowest point 0.0036 when \( n_u = 0.15 \). For any \( n_u \), the FAR always remains at 0. The Invalid ID filter results are shown in Table 6.

Before the experiment, two parameters (a number of testing segments and the number of samples in each segment) are pre-trained. Firstly, the number of testing segments is set to be 1,000. Figure 2 illustrates that the larger size of segments leads to lower FRR. However, the larger size requires longer training time. It is easy to observe that the gap is the minimum when the size of segments equals to 200. Thus we choose 200 samples as a trade-off between FRR and training time. Secondly, the experiment is repeated 5 times with the number of testing segments is 1,000, 2,000, 3,000, and 4,000, respectively. The error bars are shown in Figure 3. Finally, 1,000 test segments, each with 200 samples, are utilized to run the Invalid ID Filter Model. It should be noted that the default size of segments, 200, is the trade-off value in this paper but may not be the optimal values in practice. The training time has less effect on the running time while DeepKey works on-line.

The EEG-based filter and the combined filter are built in the same way. While the number of segments is 1,000 and the number of samples in each segment from 200 to 600, the results of three filters are listed in Table 7. Clearly, FAR is 0 for all the three filters and the FRR of the gait-based filter is hundreds of times of the FRR of the EEG-based and the combined filter, which is why we choose the gait-based filter as the Invalid ID Filter.
Table 7. Performance of gait-based, EEG-based, and combined filters. There are 1,000 testing segments and the number of samples in each segment ranges from 200 to 600.

| Filters   | Evaluation | 200  | 300  | 400  | 500  | 600  | Average |
|-----------|------------|------|------|------|------|------|---------|
| Gait-based| FAR        | 0    | 0    | 0    | 0    | 0    | 0       |
|           | FRR        | 0.0036 | 0.002 | 0    | 0    | 0    | 0.00112 |
| EEG-based | FAR        | 0    | 0    | 0    | 0    | 0    | 0       |
|           | FRR        | 0.604 | 0.66  | 0.648 | 0.6  | 0.646 | 0.6316  |
| Combined  | FAR        | 0    | 0    | 0    | 0    | 0    | 0       |
|           | FRR        | 0.605 | 0.661 | 0.648 | 0.6  | 0.646 | 0.632   |

4.5 Gait Identification Results

In this section, Gait Identification Model is evaluated by 160,000 gait samples from 8 subjects as mentioned before. The gait data are normalized before input to the model. In this specific experiment, 160,000 samples are randomly divided into 8 mini-batches and each mini-batch has 20,000 samples for test. 7 mini-batches (140,000 samples) are trained as the training set and another mini-batch (20,000 samples) is treated as the test set. All the samples are randomly allocated both in training and test sets. Every single mini-batch is in the shape of \([20000, 1, 51]\) so that the first layer of RNN structure has 51 nodes for there are 51 input features. The Gait ID (from 0 to 7) of the subject is regarded as labels. The one-hot label is deployed and each label is a list of 8 elements. The output of RNN model should have the shape of \([20000, 8]\), which means that the output layer has 8 output features and the output layer (7-th layer) has 8 nodes. Every other layer has 32 nodes and the \(\lambda\) is set as 0.001 while \(learning\ rate\) is set as 0.001. The Orthogonal Array Experiment Method, which is widely used to tuning hyper-parameters in experiment designs, is a statistically structured experiment in which several factors are applied to each experimental unit at varying levels. Range analysis on the results of experiments helps to optimize the factors and levels. All the hyper-parameters are selected by the Orthogonal Array Experiment Method [30].

Initially, the input mini-batch data have the shape of \([20000, 1, 51]\) and are reshaped to be a 2-D matrix \([20000, 51]\) for the multiplication operation. The 2-D matrix multiplies with the weight between the first and the second layer (with the weights in the shape of \([51, 32]\) where 32 is the number of nodes in the second layer), and then add the biases \([1, 32]\) in the first layer. The produced matrix with shape \([20000, 32]\) is the data in the second layer. Similarly, the data in the second layer stream to the third layer and get the data in the third layer. At last, the 7-th layer outputs the predict one-hot label with shape \([20000, 8]\).

To explore the optimal number of mini-batches, the experiment is repeated 5 times with the number of the mini-batch as 2, 4, 8, and 16. The error bars are shown in Figure 4, which shows the Gait Identification Model is stable at the accuracy of \(0.999 = 19,980/20,000\) on the classification of 20,000 samples. The confusion matrix is shown in Table 8, where the lowest precision is 0.9964 and the lowest recall is 0.9968. The definitions of Recall and Precision are given as follows:

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

where \(TP\) means true positives, \(FN\) means false negatives while \(FP\) means false positives.
Table 8. Confusion matrix of Gait ID classification. The total accuracy, the lowest precision and the lowest recall of 20,000 test samples are 0.999, 0.9964 and 0.9968, respectively.

| True label | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------|---|---|---|---|---|---|---|---|
| Predict    |   |   |   |   |   |   |   |   |
| 0          | 2467 |
| 1          | 2462 | 2 | 6 | 1 |   |   |   | 0.9964 |
| 2          | 2549 | 2 | 1 |   |   |   |   | 0.9988 |
| 3          | 2519 |   | 1 |   |   |   |   | 0.9966 |
| 4          |   | 2450 |   |   |   |   |   | 0.9992 |
| 5          | 1   |   | 2503 |   |   |   |   | 0.9996 |
| 6          |   |   |   | 2491 | 1 |   |   | 0.9992 |
| 7          |   |   |   |   | 2539 |   |   | 0.9992 |
| Total      | 2467 | 2463 | 2551 | 2523 | 2453 | 2511 | 2491 | 2541 |
| Recall     | 1   | 0.9996 | 0.9992 | 0.9984 | 0.9988 | 0.9968 | 1   | 0.9992 |

Table 9. Confusion matrix of EEG ID classification. The totally accuracy of 150 test samples is 0.993 and the lowest precision is 0.9964 and the lowest recall is 0.9968.

| True label | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------|---|---|---|---|---|---|---|---|
| Predict    |   |   |   |   |   |   |   |   |
| 0          | 18 |   |   |   |   |   |   | 1 |
| 1          | 27 |   |   |   |   |   |   | 1 |
| 2          | 19 |   |   |   |   |   |   | 1 |
| 3          |   | 14 | 1 |   |   |   |   | 0.933 |
| 4          |   | 15 |   |   |   |   |   | 1 |
| 5          |   | 24 |   |   |   |   |   | 1 |
| 6          |   | 17 |   |   |   |   |   | 1 |
| 7          |   | 15 |   |   |   |   |   | 1 |
| Total      | 18 | 27 | 19 | 14 | 15 | 24 | 17 | 16 |
| Recall     | 1   | 1   | 1   | 1   | 1   | 1   | 1   | 0.9375 |

4.6 EEG Identification Results

We set the segment window of the EEG input data as 90 samples without overlapping. Therefore, each subject (13,500 samples) has 150 = 13,500/90 segments and there are totally 1,200 segments for all the 8 subjects. The segments are labeled from 0 to 7 to indicate the order of each subject, forming a total of 1,200 labels.

After the segmentation, the data has the shape of [1200, 90, 64], that is, 1,200 segments where each segment has 90 rows and each row has 64 features. The segments will be processed by auto-regressive before being used for RNN feature extraction. The autoregressive coefficients are extracted by the `statsmodels.tsa.ar_model.AR` package. The autoregressive coefficients extraction is implemented on each column in each segment with all the parameters set as default values. The pre-processed sequence has 90 elements and the auto-regressive order is 12; therefore, totally 13 coefficients (constant coefficient is added) are produced for each column in each segment. Finally, we get the auto-regressive coefficients with shape [1200, 13, 64] and the coefficients are regarded as the input EEG data of RNN feature extraction.

\[\text{http://www.statsmodels.org/devel/generated/statsmodels.tsa.ar_model.AR.html}\]
The structure of RNN feature extraction section is described in Section 3.4. The RNN model contains 7 layers. The input layer has 64 nodes since each segment of the auto-regressive coefficients has 64 features. We use the subject ID as the training label, where the ID label (from 0 to 7) is represented as one-hot label vector. One-hot refers to a group of bits among which the legal combinations of values are only those with a single high (1) bit and all the others low (0). Therefore, the output layer (7-th layer) contains 8 nodes. The 1,200 segments from 8 subjects are randomly split into 8 mini-batch (7 mini-batch as the training dataset and 1 as the testing dataset) and each mini-batch has the shape of [150, 13, 64]. For the other 5 layers in RNN, each layer has 64 nodes. The \(\lambda\) is set as 0.004 while learning rate is set as 0.005. Similar with the parameters tuning in Section 4.5, all the hyper-parameters are selected by the Orthogonal Array Experiment Method.

The SVM classifier takes the features extracted by RNN as the input. The random state is set as 0 and other parameters set as default. We explore the performance of two classifiers: RNN and SVM, while the input data is refined by RNN structure. For a mini-batch with 150 segments, in RNN, the input EEG data (with the shape of [150, 13, 64]) flow through 7 layers RNN model and produce the ID results (with shape [150, 8]); for SVM, the input EEG data (with the shape of [150, 13, 64]) is passed through the first 6 layers of RNN structure and produce the refined feature with the shape of [150, 13, 64] (the 6-th layer has 64 nodes), and then the refined feature enters the SVM classifier and produces the ID results with the shape of [150, 1].

To explore the optimal number of mini-batches, the experiment is repeated 5 times with the number of the mini-batch set to 2, 4, 8, and 16, respectively. The error bars are shown in Figure 5. The results show the highest accuracy classification of the EEG ID from 8 subjects achieves 0.9841 when the number of mini-batches equals to 7. The SVM classifier works on 150 test segments and 149 among of them are classified correctly. The confusion matrix is shown in Table 9.

5 CONCLUSION AND FUTURE WORK

Taking the advantages of both gait- and EEG-based systems for fake-resistance, we propose Deepkey, a multimodal biometric authentication system, to overcome the limitations of traditional unimodal biometric authentication systems. This authentication system contains three independent models: an Invalid ID Filter Model, a Gait Identification Model, and an EEG Identification Model, to detect invalid gait data, recognize the Gait ID and the EEG ID, respectively. All the three models can be pre-trained efficiently off-line as the number of authorized subjects is usually limited. In particular, the Invalid ID Filter Model employs a one-class SVM to recognize invalid gait signals and achieves a FRR of 0.0036 and a FAR of 0. The Gait Identification Model adopts a 7-layer deep learning model to process gait data and classify subjects’ IDs, achieving an accuracy of 0.999. The EEG Identification Model combines three components (auto-regressive coefficients, the RNN structure, and an SVM classifier) and achieves the accuracy of 0.9841 on a public dataset. Overall, the DeepKey authentication system obtains a FAR of 0 and a FRR of 0.019.

This work sheds the light on further research on multimodal biometric authentication systems based on the gait and EEG data. Our future work will focus on enhancing the performance of the invalid ID filter component. In addition, the gait signals currently are gathered by three wearable IMUs, which may obstruct the large scale deployment in industry. Therefore, another direction is to collect gait data from non-wearable gait solutions (e.g., sensors deployed in environments).

ACKNOWLEDGMENTS

REFERENCES

[1] 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.

[2] Corey Ashby, Amit Bhatia, Francesco Tenore, and Jacob Vogelstein. 2011. Low-cost electroencephalogram (eeg) based authentication. In Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on. IEEE, 442–445.
DeepKey: An EEG and Gait Based Dual-Authentication System

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

[4] Jorge Blasco, Thomas M Chen, Juan Tapiaidor, and Pedro Peris-Lopez. 2016. A Survey of Wearable Biometric Recognition Systems. *ACM Computing Surveys (CSUR)* 49, 3 (2016), 43.

[5] Nikolaos V Boulgouris, Dimitrios Hatzinakos, and Konstantinos N Plataniotis. 2005. Gait recognition: a challenging signal processing technology for biometric identification. *IEEE Signal Processing Magazine* 22, 6 (2005), 78–90.

[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] Tianqi Chen and Carlos Guestrin. 2016. XGboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 785–794.

[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 *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 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 *Biometrics Special Interest Group (BIOSIG), 2014 International Conference of the*. IEEE, 1–4.

[12] Hasen Drira, Boulhba 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.

[13] Modern Spectral Estimation. 1988. *Theory and Application*. Bock, HG, Carra.

[14] Yu Guan, Chang-Tsun Li, and Fabio Roli. 2015. On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. *IEEE transactions on pattern analysis and machine intelligence* 37, 7 (2015), 1521–1528.

[15] Steven Goldstein. 2016. Methods and systems for voice authentication service leveraging networking. (March 8 2016). US Patent 9,282,096.

[16] Tianqi Chen and Carlos Guestrin. 2016. XGboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 785–794.
Xiang Zhang, Lina Yao, Kaixuan Chen, Xianzhi Wang, Quan Z. Sheng, and Tao Gu

[30] Xiangtao Li, Jianan Wang, and Minghao Yin. 2014. Enhancing the performance of cuckoo search algorithm using orthogonal learning method. *Neural Computing and Applications*, 24, 6 (2014), 1233–1247.

[31] 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.

[32] Dario Maio, Davide Maltoni, Raffaele Cappelli, James L Wayman, and Anil K Jain. 2002. FVC2002: Second fingerprint verification competition. In *Pattern recognition, 2002. Proceedings. 16th international conference on*. Vol. 3. IEEE, 811–814.

[33] 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.

[34] BE Manjunathswamy, Appaji M Abhishek, J Thiriveni, KR Venugopal, and LM Patnaik. 2015. Multimodal biometric authentication using ECG and fingerprint. *International Journal of Computer Applications* 111, 13 (2015).

[35] Sebastien Marcel and Jose del R Millan. 2007. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. *IEEE transactions on pattern analysis and machine intelligence* 29, 4 (2007).

[36] Sarah Mennicken, Oliver Zihler, Frida Juldaschewa, Veronika Mohnar, David Aggeler, and Elaine May Huang. 2016. It’s like living with a friendly stranger: perceptions of personality traits in a smart home. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 120–131.

[37] 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.

[38] Mark Nixon and others. 2009. Model-based gait recognition. (2009).

[39] 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.

[40] Fahreddin Sadikoglu and Selin Uzelaltinbulat. 2016. Biometric Retina Identification Based on Neural Network. *Procedia Computer Science* 102 (2016), 26–33.

[41] Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. 2004. BCI2000: a general-purpose brain-computer interface (BCI) system. *IEEE Transactions on biomedical engineering* 51, 6 (2004), 1034–1043.

[42] Torkjel Sundrøl. 2005. *Using the human gait for authentication*. Master’s thesis.

[43] Rupali L Telgad, PD Deshmukh, and Almas MN 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.

[44] Kavitha P Thomas and AP Vinod. 2016. Utilizing individual alpha frequency and delta band power in EEG based biometric recognition. In *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 004787–004791.

[45] Kavitha P Thomas and AP Vinod. 2017. EEG-Based Biometric Authentication Using Gamma Band Power During Rest State. *Circuits, Systems, and Signal Processing* (2017), 1–13.

[46] JA 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.

[47] Chen Wang, Junping Zhang, Liang Wang, Jian Pu, and Xiaoru Yuan. 2012. Human identification using temporal information preserving gait template. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 11 (2012), 2164–2176.

[48] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wfl signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 363–373.

[49] 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.

[50] Vitor Yano, Alessandro Zimmer, and Lee Luan Ling. 2012. Multimodal biometric authentication based on iris pattern and pupil light reflex. In *Pattern Recognition (ICPR), 2012 21st International Conference on*. IEEE, 2857–2860.

[51] Lina Yao, Feiping Nie, Quan Z. Sheng, Tao Gu, Xue Li, and Sen Wang. 2016. Learning from less for better: semi-supervised activity recognition via shared structure discovery. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 13–24.

[52] Sha Zhao, Julian Ramos, Jianrong Tao, Ziwen Jiang, Shijian Li, Zhaohui Wu, Gang Pan, and Anind K Dey. 2016. Discovering different kinds of smartphone users through their application usage behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 498–509.

[53] 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.