Deep Features Fusion for User Authentication Based on Human Activity

Yris Brice Wandji Piugie, Christophe Charrier, Joël Di Manno, Christophe Rosenberger

Deep Features Fusion for User Authentication Based on Human Activity

Yris Brice Wandji Piugie*†, Christophe Charrier†, Joël Di Manno* and Christophe Rosenberger†
*FIME SAS, 14000 Caen, France
†Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France
brice.wandji@fime.com, christophe.charrier@unicaen.fr, joel.dimanno@fime.com,
christophe.rosenberger@ensicaen.fr

***Abstract*—The exponential growth in the use of smartphones means that users must constantly be concerned about the security and privacy of mobile data because the loss of a mobile device could compromise personal information. To address this issue, continuous authentication systems have been proposed, in which users are monitored transparently after initial access to the smartphone. In this paper, we address the problem of user authentication by considering human activities as behavioral biometric information. We convert the behavioral biometric data (considered as time series) into a 2D color image. This transformation process keeps all the characteristics of the behavioral signal. Time series does not receive any filtering operation with this transformation and the method is reversible. This signal-to-image transformation allows us to use the 2D convolutional networks to build efficient deep feature vectors. This allows us to compare these feature vectors to the reference template vectors to compute the performance metric. We evaluate the performance of the authentication system in terms of Equal Error Rate (EER) on a benchmark UCI-HAR dataset and we show the efficiency of the approach.**

***Index Terms*—User authentication, behavioral biometrics, security, privacy, convolutional networks, human activity.**

I. Introduction

With the increasing use of smartphones to store personal and sensitive information such as bank account details, personal IDs, passwords, and credit card information, people remain constantly connected and their mobile devices are at risk of security and privacy breaches by malicious actors [1]–[3]. Traditional forms of protection such as passcodes, PINs, patterns, facial recognition, and fingerprint scans are all vulnerable to various forms of attack, including smudge attacks, side-channel attacks, and shoulder-surfing attacks [1], [3], [4].

The development of Information and Communication Technologies (ICT), as well as improvements in ambient intelligent technologies, such as sensors and smartphones, have led to the growth of smart environments [4], [5]. By using sensors, staff can save resources by recording and monitoring users or automatically reporting any unusual behavior [4], [6], [7]. For instance, in payment systems, to ensure strong customer authentication, it is necessary to implement adequate security features [1] based on authentication factors such as knowledge, possession, inherent or biometric factors [8]. Knowledge factors are based on information that the user knows, such as a password, PIN, or shared secret. Possession factors rely on an object that the user possesses, like a smart card, USB key, smartphone, or security token. Inherent or biometric factors are directly related to the user and are useful in reducing the risk of unauthorized parties discovering, disclosing, and using elements such as algorithm specifications, key length, and information entropy [9]. When Multi-Factor Authentication (MFA) is requested, using Seamless biometrics, as behavioral, improves the security without decreasing the User Experience (UX). Increasing performance of such biometrics is a high need of current industrials [3].

User authentication for logical access control, such as browsing the Internet on a laptop, is now commonly done using biometrics [3], [9], [10]. Experts employ various biometric methods, such as fingerprint, retina, and voice recognition, to design recognition systems using artificial intelligence techniques like machine learning and deep learning. Each approach has its own pros and cons, with fingerprint recognition being well-established and available in commercial products. However, these systems require input readers, such as sensors, which can vary in cost on the market [11]. Moreover, some of these biometric modalities usage are not frictionless for the subject as they have to do an additional action to authenticate.
Behavioral biometrics involves measuring a user's behavioral tendencies, which can include gait, human activity, voice recognition, signature verification, keystroke dynamics, mouse dynamics, and Graphical User Interface (GUI) usage analysis [12]. According to Bailey *et al.*, behavioral biometrics has not been as widely adopted as physiological biometrics due to the variability of the human body and mind [12]. It is worth noting that analyzing user activities does not require additional hardware.

Human activity can be one solution to enhance the security of password authentication without adding any disruptive handling for users. Industries are looking for more security without impacting too much user experience. Considered as a frictionless solution, human activity is a powerful solution to increase trust during user authentication without adding charge to the user like keystroke dynamic as a behavioral modality.

[1] http://data.europa.eu/eli/reg_del/2018/389/oj

Behavioral biometrics identification/authentication methods have lower performance compared to morphological modalities [12]. This survey aims to introduce an alternative approach using deep learning for behavioral biometrics described by time series.

The contributions of the proposed paper are numerous. The proposed research uses an image-based architecture (for a chosen behavioral biometric modality: gait analysis). A deep learning process for user authentication based on human activity is proposed. We consider only one behavioral biometrics modality which refers to the following physical activities including *laying*, *sitting*, *standing*, *walking*, *walking downstairs*, and *walking upstairs*, all of them being acquired by a smartphone. We tested many architectures for identification/authentication purposes. Generated deep features are fused through different strategies. The obtained performance on a dataset used by the research community outperforms results from the state of the art.

The paper is organized as follows. Section II contains related works on authentication systems from human activity. Section III presents the proposed method and the different tested deep learning models with the specifications and the impact of different parameters on our evaluation system. Section IV draws the experimental protocol. Section V details the experiments on benchmark datasets and the results we obtained. Section VI gives the conclusions of this work and some perspectives.

Table I: Human activity aims.

HAR tasks
Basic Activity Recognition
Daily Activity Recognition
Unusual Event Recognition
Biometric Subject Identification
Prediction of Energy Expenditures
Biometric Subject Verification/Authentication*

II. Related work

Biometrics have been widely proposed as a means of continuous user authentication in various studies [1], [15], [16], [36], [37]. In the field of continuous authentication, inertial data is used to determine the motion, orientation, and position of a device in the surrounding environment. Methods that use this type of data for non-intrusive authentication employ user behavioral features such as gait, touch screen operations, hand gestures, keyboard patterns, speech, or signature movements to generate behavioral features [1].

Zheng et al. [38] were pioneers in collecting a large dataset for continuous authentication and using a one-class distance-based classifier. They employed inertial data from the device's accelerometer and gyroscope along with touchscreen, acceleration, pressure, touch area size, and time frame information between interactions to develop user profiles of how each person held their smartphone when entering their PIN number, to identify either the genuine owner or an impostor, with an EER of up to 3.6%.

Trojahn et al. [39] also applied deep learning techniques using hand movement to authenticate smartphone users based on data collected during repeated password entry. Researchers classified users using different models such as the multilayer perceptron (MLP) [40], Bayesian Net classifiers [41], and Naïve Bayes [42].

An effective procedure for normalizing signals from smartphone accelerometers is proposed in [43] by De Marsico et al. The authors show that normalization has a positive effect on matching data from the same device, in the context of gait recognition.

Table I lists the aims of human activity. Human activity can be used for different goals (identification, authentication, soft biometrics) in different cases (continuous or static) [4], [44]. When using hand movement as a biometric solution, some systems rely on reference data, such as typing style, for verification of new samples. For identification and authentication, a reference is a specific user's typing style, while for soft biometrics, a reference is a group of users' typing style, such as male, female, or left/right-handed. This reference data is used to match or verify the identity of the user from a sample [45]. We position ourselves on the biometric verification of individuals based on human activity data.

Table II lists the state of the art on user identification and verification from human activity data. We list for each paper the method, the accuracy score, the EER value, and human activities considered for the processing. The related work in Table II focuses on activity recognition, and less work focuses on user authentication with fairly high EER values. Our goal is to verify a user based on the activities they have performed. We seek to authenticate a user based on their activities, which corresponds to an activity-based user verification approach. Therefore, our approach focuses on behavioral biometric verification of individuals using human activity data. We detail the proposed approach in the next section in order to enhance performance on user authentication with human activity data.

III. Proposed architecture

In Figure 1, we describe the proposed user authentication system based on the analysis of human activity as an extension of the work in [3]. It is composed of three different steps namely A) data collection (signal-to-image transformation), B) deep features extraction and C) verification process by scoring algorithms. We detail the following steps.

Table II: Overview of user activity identification and authentication in the state of the art.

Paper	Approach	Method	Activity	Input Source	Accuracy	EER
[13]	Action recognition	DTW	Gait	Smartphone	[83.00% – 93.00%]	[0.09% – 0.10%]
[14]	Gender recognition	SVC, RF, AdaBoost, k-NN	Looking and avoid the camera in motion	Video	[68.10% – 82.50%]	-
[15]	Continuous user authentication	Ten different classifier	Walking, sitting	Mobile devices	-	07.50%
[16]	Learning human identity from motion patterns	Dense Clockwork RNN	Walking	Smartphone	93.02%	18.17%
[17]	User identification	SVM	Pose estimation	GUI	74.35%	-
[18]	Identifying users from gait pattern	Correlation coefficients	walking	Smartphone	[72% – 88%]	7%
[19]	Gait identification using accelerometer	SVM	Walking	Mobile phone	92.7%	-
[20]	Gait recognition, analysis of approaches	SVM	Walking	Cell phone	-	33.30%
[21]	Pace independent mobile gait biometrics	Nearest neighbor	Walking	Mobile	-	7.22%
[22]	Comparison study to classify human activities	SVM, MLP, RF,Naive Bayes	Sleeping, eating, walking, falling, talking on the phone	Image	86.0%	-
[23]	Hybrid deep learning for activity and action recognition	GMM, KF, Gated Recurrent Unit	Walking, jogging, running, boxing, hand-waving, hand-clapping	Video	96.3%	-
[24]	Infer high-level rules for noninvasive ambient that help to anticipate abnormal activities	RF	Abnormal activities: agitation, alteration, screams, verbal aggression, physical aggression and inappropriate behavior	Ambient sensors	98.0%	-
[25]	Active learning to recognize human activity using Smartwatch	RF, Extra Trees, Naive Bayes, Logistic Regression, SVM	Running, walking, standing, sitting, lying down	Smartwatch	93.3%	-
[26]	Recognizing human activity using smartphone sensors	Quadratic, k-NN, ANN, SVM	Walking upstairs, downstair	Smartphone	84.4%	-
[27]	Activity recognition	CNN	Standing, sitting, laying, walking, walking downstairs, walking upstairs	Smartphone	99.30%	-
[28]	Activity recognition	Spatial Attention-aided CNN	Standing, sitting, laying, walking, walking downstairs, walking upstairs	Smartphone	99.45%	-
[29]	Action recognition	Nearest Neighbour Classifier & SVM	Bend, jack, jump, pjump, run, side, skip walk, wave	Virtual camera (6)	[90.50% – 95.70%]	-
[30]	User identification	I-vector	Gait	Mobile devices	[67.5% – 85.0%]	[06.80% – 08.90%]
[31]	Multi-view action recognition	Gaussian process + Histogram intersection kernel	Appearance of dynamic systems captured from different viewpoints	Sony AIBO robot dogs (6)	79.00%	-
[32]	Action recognition	GP-based & k-NN	golf swing (back, front, side), kicking (front, side), riding horse, run, skateboarding, swing bench, swing (side), and walk	Virtual camera	[86.90% – 88.50%]	-
[33]	User verification	HMM	25 users, 500 signatures	Samsung Galaxy Note	-	06.20%
[34]	User verification	Histogram similarity and Cycle length	Gait	Mobile devices	-	[05.00% – 09.00%]
[35]	User verification	Manhattan distance	Hand movement	Keyboard	[89.00% – 94.00%]	[06.00% – 11.00%]

A. Data collection : signal-to-image transformation

Time series analysis in the frequency domain plays an essential role in signal processing. The same is true for image analysis in the frequency domain, which plays a key role in computer vision and was even part of the standard pipeline in the early days of deep learning [46]. In this paper, we use a function that helps us transform behavioral biometric signals, which can be considered as time series, into an image, *i.e.*, we convert a signal (user activity attempt) vector v of size $1 \times m$, into a matrix M of size $n \times n$ such that: $m = n \times (n-1)/2$. If the number of features in the dataset does not meet this condition, we recommend using zero padding/ or discarding some data after preprocessing.

Such transformation is performed by using the *squareform()* [2] function in MatLab. One of the properties of the *squareform()* function is to convert a vector into a matrix, and vice versa. Conversely, the squareform of matrix M is a vector v. The *squareform()* function is bijective and this function has never been used for a transformation of human activity data.

[2]https://fr.mathworks.com/help/stats/squareform.html

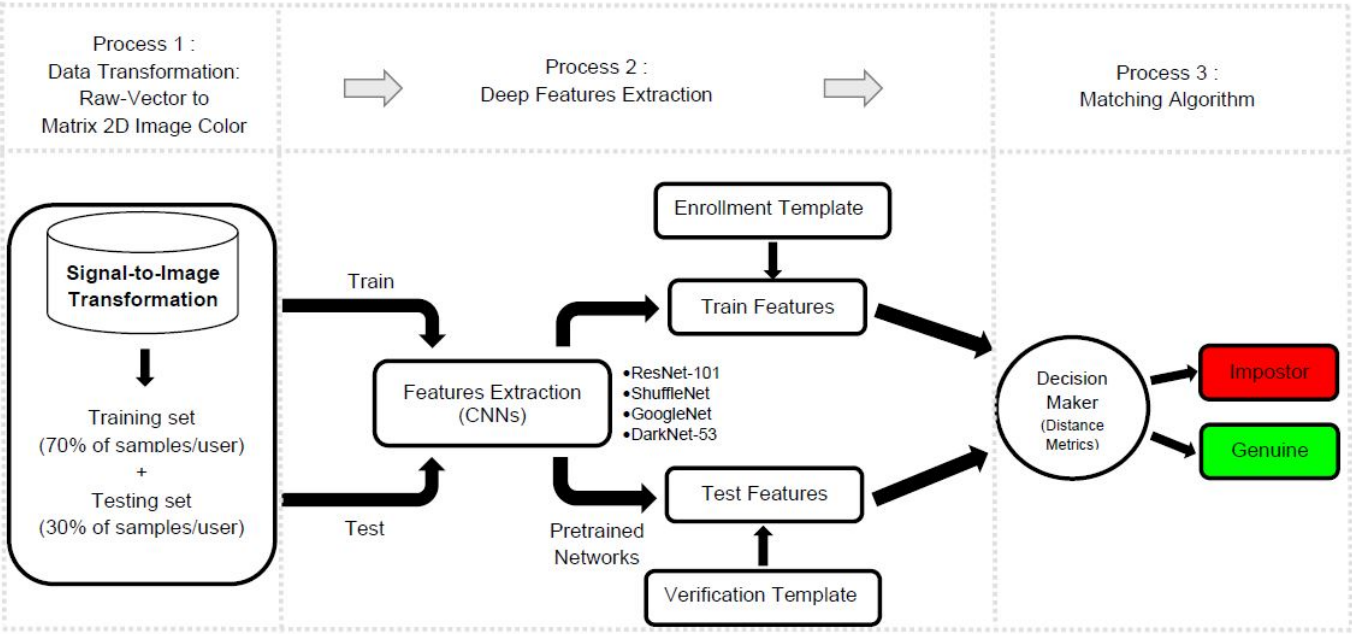

Figure 1: Architecture of the authentication system.

Table III: Architectures and optimizations hyper-parameters for the deep learning approaches

Models	#Layers	Depth	Image Input Size	Activate	Normalize	Algorithm	Loss	#Epochs	#Batch	Learning rate
ResNet-101	347	101	imresize(I, [224 224])	ReLU	Batch	SGDM	cross-entropy	300	10	0.001
ShuffleNet	172	50	imresize(I, [224 224])	ReLU	Batch	SGDM	cross-entropy	300	10	0.001
GoogleNet	144	22	imresize(I, [224 224])	ReLU	Batch	SGDM	cross-entropy	300	10	0.001
DarkNet-53	184	53	imresize(I, [256 256])	ReLU	Batch	SGDM	cross-entropy	300	10	0.001

For instance, consider the UCI-HAR dataset [3], which includes human activity information for 30 individuals. Each individual's data is represented by a raw vector of $m = 561$ features, which is then used to construct an image matrix of size 34×34, where $n = 34$. The matrix is displayed with false colors representing distance values.

We finally have a 2D image in RGB format. Figure 2 shows the step-by-step instructions to verify individual identities through our framework starting from the signal (computing the time series signal into a color image). The transformation is performed on each attempt of each user for each human activity.

B. Deep features extraction

Deep learning algorithms have been used in the last decade in several fields and are becoming more and more widespread [47], [48]. One advantage of using such approaches relies in its capabilities to provide relevant features at deeper layers which can be used as feature vectors by any dissimilarity measure. In this work, we generate deep features vectors by transfer learning from four different deep networks namely ResNet-101, ShuffleNet, GoogleNet and DarkNet-53. In the literature, these models were firstly pre-trained on the ImageNet dataset [4] and they are the most recent successful deep learning architectures for image classification [49] and can be used for an authentication context since authentication can be considered as the result of a binary classification problem (genuine or impostor).

Table III summarizes the architecture and the optimization hyper-parameters for the four trial deep networks where the network depth is defined as the largest number of sequential convolutional or fully connected layers on a path from the input layer to the output layer. As input, networks take RGB images format describing a human activity. These convolutional networks are used to build an output feature vector (by extraction at the last layer of the convolutional network), which is then compared to the reference/test model.

C. Matching algorithms

Deep architectures as previously explained generate feature vectors that can be used as reference/test templates. We need a matching algorithm to compare them and make the verification decision. Many distance metrics can be used to compute a distance score [9] between a reference (x_s) and a sample (x_t) such as:

- The Minkowski distance

$$d = \sum_{j=1}^{n} |x_{sj} - x'_{tj}| \quad (1)$$

[3] http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

[4] https://image-net.org

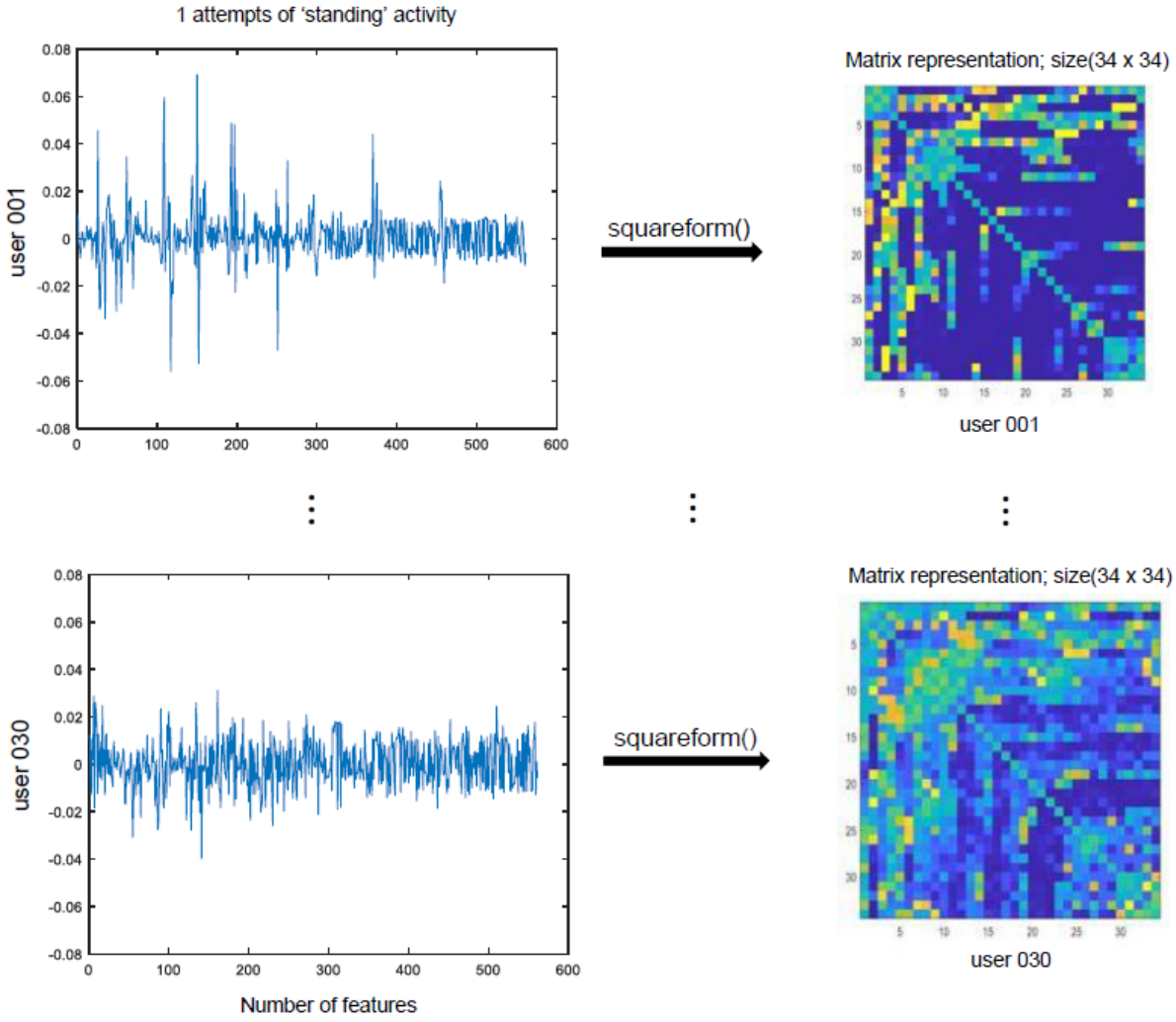

Figure 2: Examples of the obtained results when the signal-to-image transformation is applied.

Table IV: Activities, sample number of each activity and their descriptions on UCI-HAR dataset [26].

Activity	Abbreviation	No. of Samples	Sample Percent of Each Human Activity	Description
Laying	lyx	1722	16.72%	Subject sleeps or lies down on a bed
Sitting	six	1544	14.99%	Subject sits on a chair either working or resting
Standing	stx	1406	13.65%	Subject stands and talks to someone
Walking	wlx	1777	17.25%	Subject goes down multiple flights
Walking Downstair	wdn	1906	18.51%	Subject goes down multiple flights
Walking Upstairs	wup	1944	18.88%	Subject goes up multiple flights

- The Euclidean distance

$$d^2 = (x_s - x_t)(x_s - x_t)' \tag{2}$$

- The Cosine distance

$$d = 1 - \frac{x_s x_t'}{\sqrt{(x_s x_s')(x_t x_t')}} \tag{3}$$

Once we obtain a matching score, we decide if the user is authenticated by a simple thresholding approach (accept when the score is upper a given threshold).

Algorithm 1 Scores Computation

Input Output data, distance, N, M INTRA, INTER
Scores_computation data, distance, N, M
Initialize counters `cptra = 1 cpter = 1`
Compute intraclass and interclass scores
$i = 1\ N\ l = 1\ N\ j = 2\ M\ i == l$

```
INTRA(cptra) = pdist([data(M*(i-1)+1,:);
data(M*(l-1)+j,:)], distance)
cptra = cptra + 1
INTER(cpter) = pdist([data(M*(i-1)+1,:);
data(M*(l-1)+j,:)], distance)
cpter = cpter + 1
```

Return INTRA, INTER

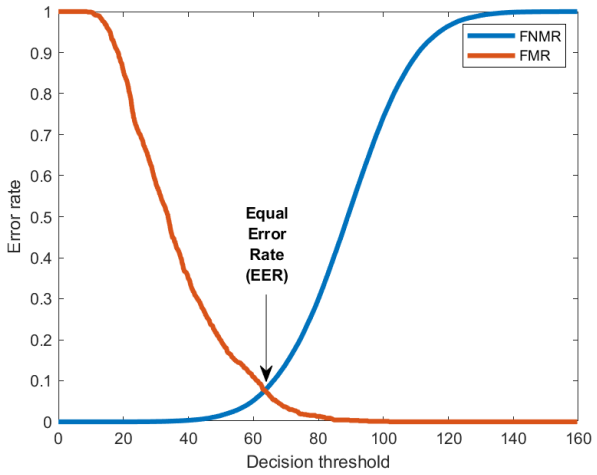

Figure 3: Relationship between FMR, FNMR and EER (source [3]).

Algorithm 1 describes the procedure for computing INTRA scores for similarity and INTER scores for dissimilarity. These scores are used to compute the FMR, FNMR and EER scores later on.

IV. Experimental protocol

We draw in this part the experimental protocol we follow in this work. We detail the used biometric dataset and the performance metrics.

A. Database

We use in this work the UCI-HAR database [26] which was collected with data from 30 people aged between 19 and 48 years. Each person performed 6 physical activities such as *sitting*, *standing*, *laying walking*, *walking upstairs* and *walking downstairs*. The data were collected from a *Samsung Galaxy S II* mobile phone handset using the accelerometer and gyroscope (3-axial raw signals with *tAcc-XYZ* and *tGyro-XYZ*) sensors at a frequency of 50Hz. The collection was obtained with the smartphone located at the user's waist. All steps of data collection were recorded and the data was manually labeled. UCI-HAR contains $10,299$ samples.

Table IV presents the activities, the abbreviation of each activity, the proportion of activity samples and their descriptions. For each signal of each activity, the signal-to-image transformation (as mentioned in section III) is applied to obtain a 2D color image. To the best of our knowledge, such transformation with the *squareform()* function applied to the UCI-HAR dataset does not exist in the literature up to now. Among the transformed samples of each user, 70% out of 100% samples (attempts per subject) are used for the training set and the remaining 30% for the testing set.

B. Pretrained models

As previously mentioned, the used pretrained models for comparison of the different architectures are the following networks: ResNet-101, ShuffleNet, GoogleNet, and DarkNet-53. It is not the purpose of this paper to provide theoretical information on how each of these architectures work; more details on each of them can be found in [3].

C. Performance metrics

The authentication/verification stage involves acquiring and processing raw data to create a biometric template, which is then compared to reference templates in the dataset. A matching algorithm is used to determine the similarity between the biometric sample and existing reference templates. Scores are calculated based on features extracted from deep networks, and three distance metrics described in subsection III-C are applied to evaluate the degree of similarity between the activity of each user.

Two important error rates are used to assess the performance of a biometric authentication system according to ISO19795 [50]: 1) False Match Rate (FMR) and 2) False Non-Match Rate (FNMR).

1) The FMR is the proportion of a specified set of completed non-mated comparison trials that result in a comparison decision of *match*,
2) The FNMR is the proportion of completed mated comparison trials that result in a comparison decision of *non-match*.

The Equal Error Rate (EER) is obtained when the biometric decision threshold is set to have the FMR value equal to the FNMR one as depicted in Figure 3. It can be seen as a compromise between usability and security. The goal of a matching algorithm is to minimize this value. The lower the value of EER, the better the performance of the authentication system is. This error rate is the most commonly used in the literature to evaluate the performance of biometric systems. In this work, we evaluate the performance of the proposed architecture in terms of EER. Experiments has been realized on an Intel Core i5-9600K CPU 3.70 GHz computer equipped with 16.00 GB of RAM with MatLab.

V. Results and Discussion

In this section, we present the experimental results we obtained. We tried to structure them by addressing some questions concerning the performance of the proposed method on such behavioral biometric datasets. Note that we considered 70% of user samples (attempts per subject) for the learning phase and 30% for the testing one.

A. Which performance can we expected on a larger dataset ?

First, we consider all the six activity sub-datasets defined as the UCI-HAR dataset (**Fusion of features**). Table V draws the obtained results for the user verification task considering

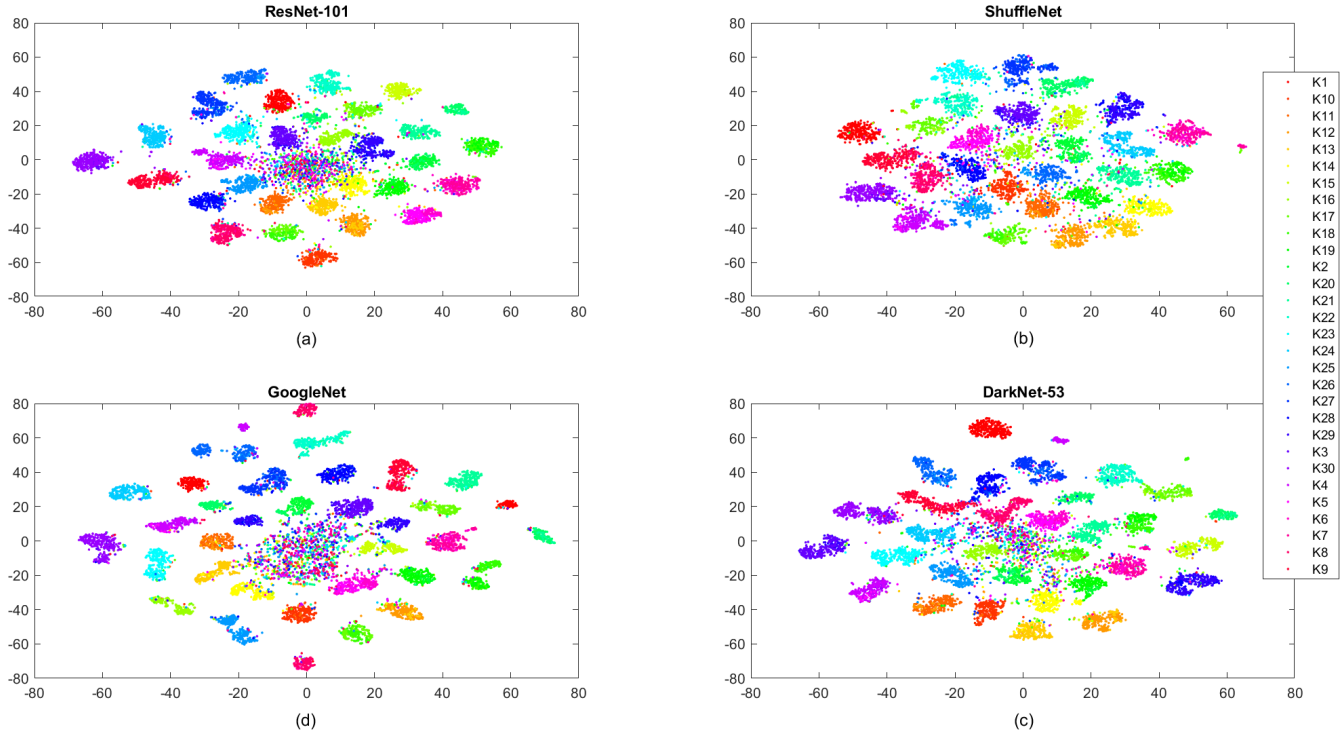

Figure 4: Visual inspection of deep features projection from (a) ResNet-101, (b) ShuffleNet, (c) DarkNet-53 and (d) GoogleNet.

the three distances. One can observe that whatever the deep architecture, the distance minimizing the EER value is the Cosine one. In the rest of this work, all results are given considering this distance.

Table V: EER value on HAR dataset for the three tested distances.

Models	$EER_{mananthan}$	$EER_{euclidean}$	EER_{cosine}
ResNet-101	22.69%	17.71%	**12.48%**
ShuffleNet	14.77%	14.63%	**11.57%**
GoogleNet	14.88%	14.56%	**13.52%**
DarkNet-53	17.46%	14.31%	**11.72%**

We visually inspected our four deep networks by performing a feature projection through the T-SNE (t-Distributed Stochastic Neighbor Embedding) function as shown in Figure 4 for each architecture. We observe that the deep features projection form a nearly distant cluster in ShuffleNet ($EER = 11.57\%$) than DarkNet-53 ($EER = 11.72\%$), ResNet-101 ($EER = 12.48\%$) and GoogleNet ($EER = 13.52\%$). This result is correlated with the fact that ShuffleNet performs better than other networks in terms of EER value.

B. How well can we perform on each activity separately?

In this section, we consider each activity separately as shown in Table IV to generate six sub-datasets among others laying (1722 samples for the 30 subjects), sitting (1544 samples for the 30 subjects), standing (1406 samples for the 30 subjects), walking (1777 samples for the 30 subjects), walking downstairs (1906 samples for the 30 subjects), walking upstairs (1944 samples for the 30 subjects). We illustrate the four architectures (namely ResNet-101, ShuffleNet and GoogleNet and DarkNet-53) and we draw the model on each sub-datasets separately.

This is illustrated by the block 1 in Figure 5. The best model among the four deep networks for each activity in terms of EER value are: standing (GoogleNet=13.52%), sitting (GoogleNet=15.15%), laying (GoogleNet=07.78%), walking (ShuffleNet=07.02%), walking downstairs (ShuffleNet=08.14%) and walking upstairs (GoogleNet=06.88%). Here, we try to verify one user among the 30 users based on their activities separately.

We note that we do not have the same performance from one activity to another. So, using 70% of samples (attempts per subject) for the generation of the reference template does not provide exceptional results (with an EER value between 06.88% to 19.65%) as shown in the block 1 in Figure 5. Obviously, if we had more samples per subject (or by combining activities), we could expect to obtain a better performance.

C. What performance can be achieved if the user performs more than one activity ?

In this part, it is assumed that a person achieved more than one activity to authenticate himself/herself. We merge by summing the legitimate and impostor scores considering the number of samples (activity attempt) per user (**Fusion of scores**). Table VI shows the obtained results if we used all the six activities (*i.e.*, simulating a user achieving 6 activities to be authenticated). ShuffleNet comes out as the best method with an EER score of 03.58% as presented in Table VI.

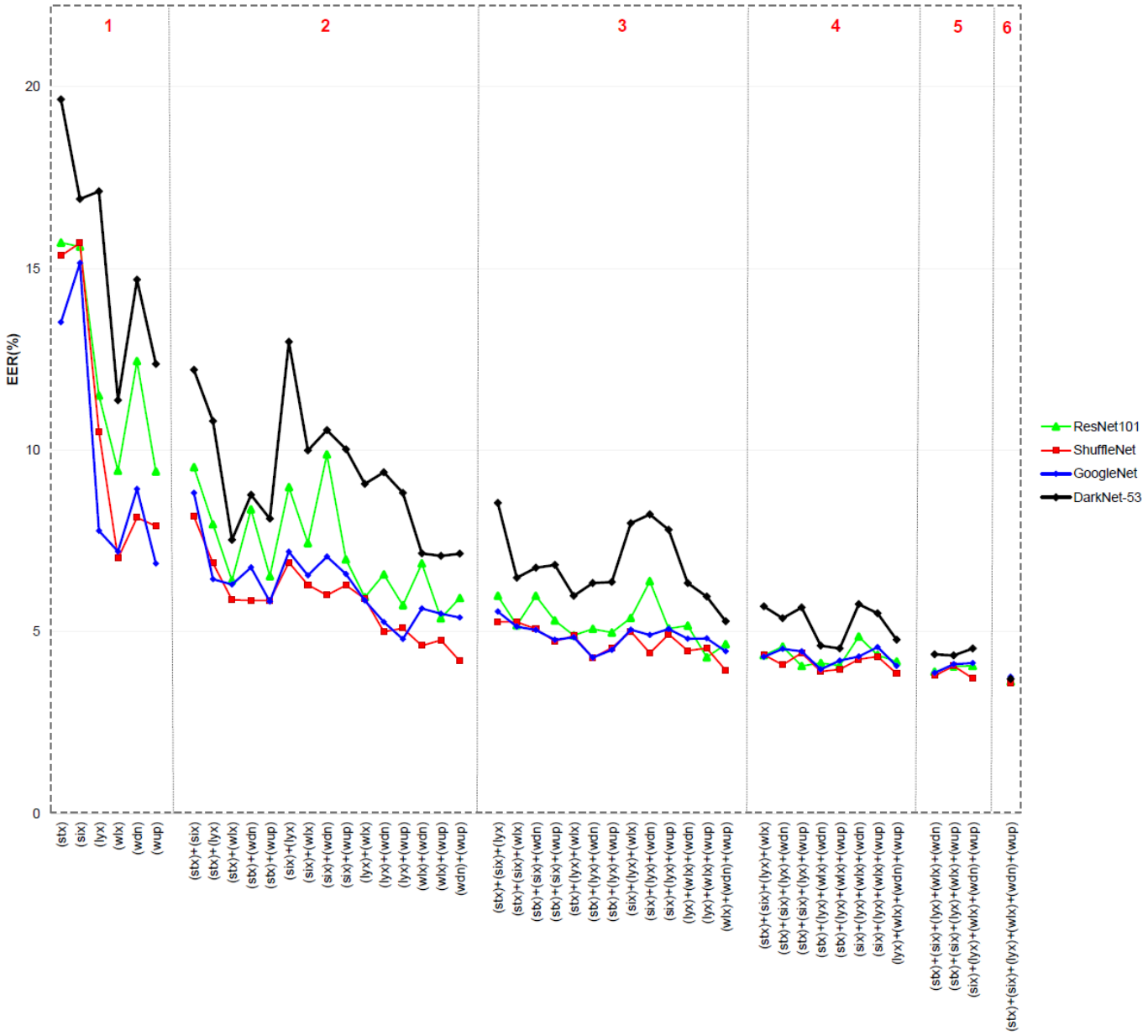

Figure 5: EER rate on deep architectures for the multi-instance biometric system. In block 1, we have (stx), (six), (lyx), (wlx), (wdn) and (wup) activities. In block 2, we have the fusion of inter and intra class score from {(stx)+(six)} to {(wdn)+(wup)} activities respectively. In block 3, we have {(stx)+(six)+(lyx)} to {(wlx)+(wdn)+(wup)}. In block 4, {(stx)+(six)+(lyx)+(wlx)} to {(lyx)+(wlx)+(wdn)+(wup)}, in Block 5, {(stx)+(six)+(lyx)+(wlx)+(wdn)} to {(six)+(lyx)+(wlx)+(wdn)+(wup)} and in Block 6 {(stx)+(six)+(lyx)+(wlx)+(wdn)+(wup)}.

ShuffleNet is ahead of ResNet-101 (03.63%), GoogleNet (03.76%) and DarkNet-53 (03.70%).

Table VI: Performance evaluation on the multi-instance biometric system by fusion of features and scores level on UCI-HAR dataset.

Models (EER_{cosine})	**Fusion of features**	**Fusion of scores**
ResNet-101	12.48%	3.63%
ShuffleNet	**11.57%**	**3.58%**
GoogleNet	13.52%	3.76%
DarkNet-53	11.72%	3.70%

To complete these results, we studied the obtained performance versus the number of activities (*laying*, *sitting*, *standing*, *walking*, *walking downstairs* and *walking upstairs*) achieved by a user in a multi-instance context. Figure 5 highlights the EER value obtained for each case.

- If we use 2 activities, we obtain an EER value between $[04.20\% - 12.98\%]$ illustrated by block 2 in Figure 5.
- If we have 3 activities, we have an EER value between $[03.94\% - 08.55\%]$ represented in block 3.

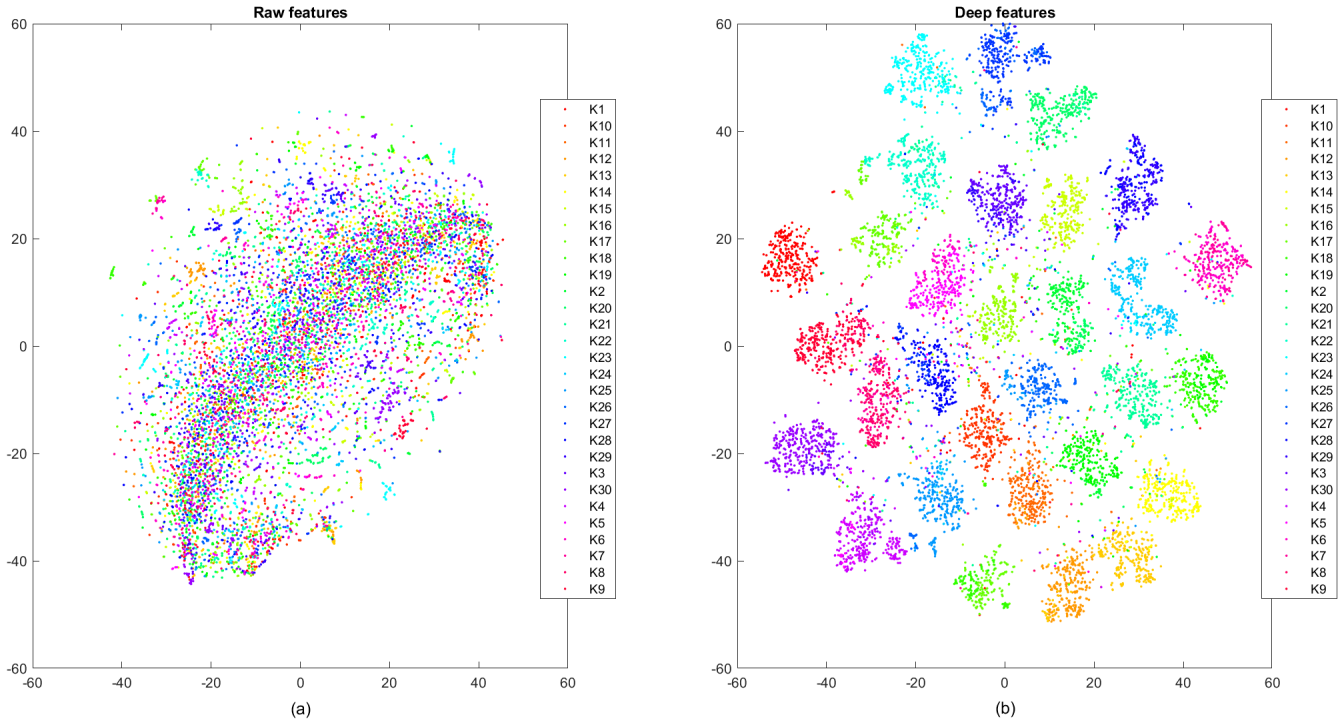

Figure 6: t-SNE projection of (a) raw features and (b) deep features extracted from the top-performing method (ShuffleNet). The x-axis corresponds to dimension 1, while the y-axis corresponds to dimension 2.

- If we use 4 activities, we have an EER value around $[03.85\% - 05.71\%]$ depicted by block 4.
- If we use 5 activities, we have an EER value around $[03.72\% - 04.54\%]$ shown by block 5.
- If we use 6 activities (*laying* + *sitting* + *standing* + *walking* + *walking downstairs* + *walking upstairs*), we get an EER value around $[03.58\% - 03.76\%]$ depicted by block 5. This shows we can decrease easily the EER value for this kind of authentication.

We find that the value of EER obtained by fusion of the scores of each activity decreases for all architectures. It also appears from this work that the more information we have, the better the performance can be, this is not surprising. With a more extensive database, we could expect to get better results (*i.e.* with an EER value very close to 0%) by increasing the number of samples per user [3].

D. Discussion

Due to their ability to perform sensitive operations like mobile banking, communication, and personal data storage, smartphones have become a crucial part of daily life. This has led to a greater need for secure authentication methods to protect critical information from unauthorized access. [1].

The purpose of this work is to analyze several information from user activity in order to authenticate himself/herself. A comparative analysis of the four architectures on UCI-HAR dataset allows us to identify the best performance for a continuous authentication. From Table V, it is provided by the ShuffleNet architecture with an EER value equal to 11.57%. Figure 6 shows a visual inspection of features (raw versus deep features) projection for ShuffleNet. It shows clearly the good separability of the deep features. Multi-instance systems intend to capture samples of two or more different instances of the same biometric characteristics. The Table VI shows that for an authentication performed on human activity, the best verification scores are obtained on the fusion of scores (EER = 03.58%) as opposed to the fusion of features (EER = 11.57%) on ShuffleNet among the four different deep neural network architectures.

In the literature, several works notably [51]–[58] shown in Table II and others in Table VII have been carried out only on the recognition of activities (where the target is : standing, sitting, laying, walking, walking downstairs, walking upstairs) from the UCI-HAR dataset. The used methods are respectively Deep CNN-LSTM with Self-Attention (accuracy = 93.11%), linear SVC (accuracy = 96.50%), LSTM-CNN (accuracy = 95.78%), SVM (accuracy = 97.12%), Lego-CNN (accuracy = 96.90%), LSTM (accuracy = 97.40%), CNN (accuracy = 96.40%). Among these works, there are several studies that convert the 1D time-series into a 2D image representation and then apply 2D image-based feature extraction technique [27]–[29]. These transformations are not reversible and the related works are typically based on activity or action recognition. However, this work focuses on activity-based user verification.

We can compare our results with research works that has been performed on the UCI-HAR dataset in Table VIII.

Table VII: Comparison with other published works (target = activities).

Dataset	Author/S (ref)	Years	Classifiers	Accuracy	EER
UCI-HAR (target = activities)	Sanchez *et al.* [27]	2022	CNN	99.30%	-
UCI-HAR (target = activities)	Sarkar *et al.* [28]	2022	Spatial Attention-aided CNN	99.45%	-
CMU mocap (target = actions)	Junejo *et al.* [29]	2008	Nearest Neighbour Classifier & SVM	[90.50% – 95.70%]	-
Naturalistic McGill University gait dataset and Osaka University gait dataset	Zhong *et al.* [30]	2014	I-vector	[67.5% – 85.0%]	[06.80% – 08.90%]
IXMAS (target = actions)	Korner *et al.* [31]	2013	Gaussian process + Histogram intersection kernel	79.00%	-
UCF sports (target = actions)	Chuan *et al.* [32]	2015	GP-based & k-NN	[86.90% – 88.50%]	-

Table VIII: Comparison with other published works on user activity (target = users).

Dataset	Author/S (ref)	Years	Classifiers	Accuracy	EER
UCI-HAR (target = users)	**This Paper**	**2022**	**ShuffleNet**	-	**03.57%**
UCI-HAR (target = users)	Mekruksavanich *et al.* [1]	2021	DeepConvLSTM	-	5.10%
Touch gestures data	Patel *et al.* [15]	2016	Ten classifiers	-	07.50%
WISDM	Zhang *et al.* [16]	2019	Dense Clockwork RNN	-	18.17%
Gait signal data	Mantyjarvi *et al.* [18]	2005	Correlation coefficients	-	07%
Biometric gait data	Muaazz *et al.* [20]	2013	SVM	-	33.30%
Mobile gait data	Zhong *et al.* [21]	2015	Nearest neighbor	-	07.22%

Mekruksavanich *et al.* [1] in 2021 work on deep learning approaches for continuous authentication based on activity patterns using mobile sensing. They had obtained for each distinct activity an EER score 5.10% with the DeepConvLSTM network. By merging the legitimate and impostor scores of each activity, we obtain an EER score of 03.58% with the ShuffleNet network. This means that during a verification scheme, the more activities a user performs, the better it can be authenticated by our framework. To the best of our knowledge, in the literature, there is no work addressing fusion of scores on the basis of the UCI-HAR dataset.

VI. Conclusion and perspectives

This survey is based on a new method for user authentication by analyzing human activities in this paper. We tested various deep learning classifiers (ResNet-101, ShuffleNet, GoogleNet, and DarkNet-53) on the UCI-HAR benchmark dataset for authentication applications. The results showed that using a combination of motion sensor data resulted in the lowest Equal Error Rate (EER) for binary classification.

The aim of this study was to determine the effectiveness of deep learning architectures in authenticating smartphone users based on their physical activity patterns measured by the accelerometer, gyroscope, and magnetometer sensors on their smartphones. We demonstrated that our new framework outperforms current state-of-the-art methods in terms of Equal Error Rate (EER) for continuous smartphone authentication utilizing various sensor data. The main contribution of this paper is to answer how well deep learning approaches could verify individual identities by using smartphone sensing data from 30 users. Since in the state of the art, results are given for the activity classification, our second contribution is the use of another signal-to-image transformation (a bijective transformation) of the input data, which leads to improved authentication results.

For future research, we plan to study how to improve the security of biometric continuous authentication systems linked to human activities by creating innovative Presentation Attack Instruments for laboratory evaluations and synthetically generated Human Activities Databases. We intend to consider the quality assessment of human activities with deep learning architectures in order to enhance the authentication results.

Acknowledgment

The present manuscript builds upon the previously published work entitled "Keystroke Dynamics based User Authentication using Deep Learning Neural Networks." It was originally presented at the CYBERWORLDS 2022 international conference and received recognition as the "Best Paper Award" during the event. The authors would like to thank the “Normandy Region”, FIME SAS, and the ANRT for their financial support of this work.

References

[1] S. Mekruksavanich and A. Jitpattanakul, “Deep learning approaches for continuous authentication based on activity patterns using mobile sensing,” *Sensors*, vol. 21, no. 22, p. 7519, 2021.

[2] C. Nugier, D. Leblanc-Albarel, A. Blaise, S. Masson, P. Huynh, and Y. B. W. Piugie, “An upcycling tokenization method for credit card numbers,” in *SECRYPT 2021-18th International Conference on Security and Cryptography*, 2021.

[3] Y. B. W. Piugie, J. Di Manno, C. Rosenberger, and C. Charrier, “Keystroke dynamics based user authentication using deep learning neural networks,” in *2022 INTERNATIONAL CONFERENCE ON CYBERWORLDS (CW 2022)*, 2022.

[4] Y. B. W. Piugie, J. Manno, C. Rosenberger, and C. Charrier, “How artificial intelligence can be used for behavioral identification?” in *2021 International Conference on Cyberworlds (CW)*, 2021.

[5] A. Visvizi, J. Jussila, M. D. Lytras, and M. Ijäs, “Tweeting and mining oecd-related microcontent in the post-truth era: a cloud-based app,” *Computers in Human Behavior*, vol. 107, p. 105958, 2020.

[6] A. Rasekh, C.-A. Chen, and Y. Lu, “Human activity recognition using smartphone,” *arXiv preprint arXiv:1401.8212*, 2014.

[7] Y. B. W. Piugie, D. Tchiotsop, A. N. K. Telem, and E. B. M. Ngouonkadi, "Denoising of electroencephalographic signals by canonical correlation analysis and by second-order blind source separation," in *2019 IEEE AFRICON*. IEEE, 2019, pp. 1–8.

[8] E. Cherrier, "Authentification biométrique: comment (ré) concilier sécurité, utilisabilité et respect de la vie privée?" Ph.D. dissertation, Normandie Université, 2021.

[9] D. Migdal, "Contributions to keystroke dynamics for privacy and security on the internet," Ph.D. dissertation, Normandie Université, 2019.

[10] M. Yohan, H. Hanry, and D. Dion, "Keystroke dynamic classification using machine learning forpassword authorization," in *3rd International Conference on Computer Science and Computational Intelligence, Procedia Computer Science*, 2018.

[11] D. I. Kim, S. Lee, and J. S. Shin, "A new feature scoring method in keystroke dynamics-based user authentications," *IEEE Access*, vol. 8, pp. 27 901–27 914, 2020.

[12] K. O. Bailey, J. S. Okolica, and G. L. Peterson, "User identification and authentication using multi-modal behavioral biometrics," *Computers & Security*, vol. 43, pp. 77–89, 2014.

[13] M. D. Marsico and A. Mecca, "Biometric walk recognizer: gait recognition by a single smartphone accelerometer," *Multimedia Tools and Applications*, vol. 76, pp. 4713–4745, 2017.

[14] P. Barra, C. Bisogni, M. Nappi, D. Freire-Obregón, and M. Castrillón-Santana, "Gait analysis for gender classification in forensics," in *Dependability in Sensor, Cloud, and Big Data Systems and Applications: 5th International Conference, DependSys 2019, Guangzhou, China, November 12–15, 2019, Proceedings 5*. Springer, 2019, pp. 180–190.

[15] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 49–61, 2016.

[16] M. Zhang, "Gait activity authentication using lstm neural networks with smartphone sensors," in *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*. IEEE, 2019, pp. 456–461.

[17] G. Gao, Y. Yu, J. Yang, G.-J. Qi, and M. Yang, "Hierarchical deep cnn feature set-based representation learning for robust cross-resolution face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[18] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S.-M. Makela, and H. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 2. IEEE, 2005, pp. ii–973.

[19] H. M. Thang, V. Q. Viet, N. D. Thuc, and D. Choi, "Gait identification using accelerometer on mobile phone," in *2012 International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, 2012, pp. 344–348.

[20] M. Muaaz and R. Mayrhofer, "An analysis of different approaches to gait recognition using cell phone based accelerometers," in *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*, 2013, pp. 293–300.

[21] Y. Zhong, Y. Deng, and G. Meltzner, "Pace independent mobile gait biometrics," in *2015 IEEE 7th international conference on biometrics theory, applications and systems (BTAS)*. IEEE, 2015, pp. 1–8.

[22] P. M. D. Alex, A. Ravikumar, J. Selvaraj, and A. Sahayadhas, "Research on human activity identification based on image processing and artificial intelligence," *Int. J. Eng. Technol*, vol. 7, 2018.

[23] N. Jaouedi, N. Boujnah, and M. S. Bouhlel, "A new hybrid deep learning model for human action recognition," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 4, pp. 447–453, 2020.

[24] M. Á. Antón, J. Ordieres-Meré, U. Saralegui, and S. Sun, "Non-invasive ambient intelligence in real life: Dealing with noisy patterns to help older people," *Sensors*, vol. 19, no. 14, p. 3113, 2019.

[25] F. Shahmohammadi, A. Hosseini, C. E. King, and M. Sarrafzadeh, "Smartwatch based activity recognition using active learning," in *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 2017, pp. 321–329.

[26] D. Anguita, A. Ghio, L. Oneto, X. Parra Perez, and J. L. Reyes Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning*, 2013, pp. 437–442.

[27] A. Sanchez Guinea, M. Sarabchian, and M. Mühlhäuser, "Improving wearable-based activity recognition using image representations," *Sensors*, vol. 22, no. 5, p. 1840, 2022.

[28] A. Sarkar, S. Hossain, and R. Sarkar, "Human activity recognition from sensor data using spatial attention-aided cnn with genetic algorithm," *Neural Computing and Applications*, pp. 1–27, 2022.

[29] I. N. Junejo, E. Dexter, I. Laptev, and P. Púrez, "Cross-view action recognition from temporal self-similarities," in *European Conference on Computer Vision*. Springer, 2008, pp. 293–306.

[30] Y. Zhong and Y. Deng, "Sensor orientation invariant mobile gait biometrics," in *IEEE international joint conference on biometrics*. IEEE, 2014, pp. 1–8.

[31] M. Körner and J. Denzler, "Temporal self-similarity for appearance-based action recognition in multi-view setups," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2013, pp. 163–171.

[32] C. Sun, I. N. Junejo, M. Tappen, and H. Foroosh, "Exploring sparseness and self-similarity for action recognition," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2488–2501, 2015.

[33] F. J. Zareen and S. Jabin, "Authentic mobile-biometric signature verification system," *IET Biometrics*, vol. 5, no. 1, pp. 13–19, 2016.

[34] D. Gafurov, K. Helkala, and T. Søndrol, "Biometric gait authentication using accelerometer sensor." *J. comput.*, vol. 1, no. 7, pp. 51–59, 2006.

[35] S. Parkinson, S. Khan, A. Crampton, Q. Xu, W. Xie, N. Liu, and K. Dakin, "Password policy characteristics and keystroke biometric authentication," *IET Biometrics*, vol. 10, no. 2, pp. 163–178, 2021.

[36] Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani, "Hmog: New behavioral biometric features for continuous authentication of smartphone users," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 5, pp. 877–892, 2015.

[37] G. Giorgi, A. Saracino, and F. Martinelli, "Using recurrent neural networks for continuous authentication through gait analysis," *Pattern Recognition Letters*, vol. 147, pp. 157–163, 2021.

[38] N. Zheng, K. Bai, H. Huang, and H. Wang, "You are how you touch: User verification on smartphones via tapping behaviors," in *2014 IEEE 22nd International Conference on Network Protocols*. IEEE, 2014, pp. 221–232.

[39] M. Trojahn and F. Ortmeier, "Toward mobile authentication with keystroke dynamics on mobile phones and tablets," in *2013 27th International Conference on Advanced Information Networking and Applications Workshops*. IEEE, 2013, pp. 697–702.

[40] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, classifiaction," 1992.

[41] R. Kohavi *et al.*, "Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid." in *Kdd*, vol. 96, 1996, pp. 202–207.

[42] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello, and G. Taylor, "Learning human identity from motion patterns," *IEEE Access*, vol. 4, pp. 1810–1820, 2016.

[43] M. De Marsico, D. De Pasquale, and A. Mecca, "Embedded accelerometer signal normalization for cross-device gait recognition," in *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2016, pp. 1–5.

[44] Z. Chen, H. Cai, L. Jiang, W. Zou, W. Zhu, and X. Fei, "Keystroke dynamics based user authentication and its application in online examination," in *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 2021, pp. 649–654.

[45] D. Migdal, "Contributions to keystroke dynamics for privacy and security on the Internet," Theses, Normandie Université, Nov. 2019. [Online]. Available: https://tel.archives-ouvertes.fr/tel-02518436

[46] C. Vasconcelos, H. Larochelle, V. Dumoulin, R. Romijnders, N. L. Roux, and R. Goroshin, "Impact of aliasing on generalization in deep convolutional networks," *arXiv preprint arXiv:2108.03489*, 2021.

[47] S. Maheshwary, S. Ganguly, and V. Pudi, "Deep secure: A fast and simple neural network based approach for user authentication and identification via keystroke dynamics," in *IWAISe: First International Workshop on Artificial Intelligence in Security*, vol. 59, 2017.

[48] L. Aversano, M. L. Bernardi, M. Cimitile, and R. Pecori, "Continuous authentication using deep neural networks ensemble on keystroke dynamics," *PeerJ Computer Science*, 2021.

[49] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[50] ISO, "Information technology — Biometric performance testing and reporting — part 1: Principles and framework," International Organization for Standardization, Geneva, CH, Standard ISO/IEC 19795-1:2021, 2021.

[51] M. A. Khatun, M. A. Yousuf, S. Ahmed, M. Z. Uddin, S. A. Alyami, S. Al-Ashhab, H. F. Akhdar, A. Khan, A. Azad, and M. A. Moni, "Deep cnn-lstm with self-attention model for human activity recognition using wearable sensor," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–16, 2022.

[52] S. Khare, S. Sarkar, and M. Totaro, "Comparison of sensor-based datasets for human activity recognition in wearable iot," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*. IEEE, 2020, pp. 1–6.

[53] T. Mahmud, A. S. Sayyed, S. A. Fattah, and S.-Y. Kung, "A novel multi-stage training approach for human activity recognition from multimodal wearable sensor data using deep neural network," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1715–1726, 2020.

[54] K. Xia, J. Huang, and H. Wang, "Lstm-cnn architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56 855–56 866, 2020.

[55] A. Jain and V. Kanhangad, "Human activity classification in smartphones using accelerometer and gyroscope sensors," *IEEE Sensors Journal*, vol. 18, no. 3, pp. 1169–1177, 2017.

[56] Y. Tang, Q. Teng, L. Zhang, F. Min, and J. He, "Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors," *IEEE Sensors Journal*, vol. 21, no. 1, pp. 581–592, 2020.

[57] N. Tufek, M. Yalcin, M. Altintas, F. Kalaoglu, Y. Li, and S. K. Bahadir, "Human action recognition using deep learning methods on limited sensory data," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3101–3112, 2019.

[58] T. Zebin, P. J. Scully, N. Peek, A. J. Casson, and K. B. Ozanyan, "Design and implementation of a convolutional neural network on an edge computing smartphone for human activity recognition," *IEEE Access*, vol. 7, pp. 133 509–133 520, 2019.