



**HAL**  
open science

## Deep features fusion for user authentication based on human activity

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

► **To cite this version:**

Yris Brice Wandji Piugie, Christophe Charrier, Joël Di Manno, Christophe Rosenberger. Deep features fusion for user authentication based on human activity. *IET Biometrics*, 2023, 12 (4), pp.222-234. 10.1049/bme2.12115 . hal-04379327

**HAL Id: hal-04379327**

**<https://hal.science/hal-04379327>**

Submitted on 27 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## ORIGINAL RESEARCH

# Deep features fusion for user authentication based on human activity

Yris Brice Wandji Piugie<sup>1,2</sup>  | Christophe Charrier<sup>2</sup> | Joël Di Manno<sup>1</sup> |  
Christophe Rosenberger<sup>2</sup>

<sup>1</sup>FIME SAS, Caen, France

<sup>2</sup>Normandie University, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

## Correspondence

Yris Brice Wandji Piugie.  
Email: [brice.wandji@fime.com](mailto:brice.wandji@fime.com)

## Funding information

Association Nationale de la Recherche et de la Technologie; Région Normandie; FIME SAS

## 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 study, the authors address the problem of user authentication by considering human activities as behavioural biometric information. The authors convert the behavioural biometric data (considered as time series) into a 2D colour image. This transformation process keeps all the characteristics of the behavioural 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 them to compare these feature vectors to the reference template vectors to compute the performance metric. The authors evaluate the performance of the authentication system in terms of Equal Error Rate on a benchmark University of Californy, Irvine Human Activity Recognition dataset, and they show the efficiency of the approach.

## KEYWORDS

behavioural biometrics, biometric system performance evaluation, convolutional neural nets, feature extraction, gait analysis, gait biometrics, image classification, performance evaluation, security, signal representation

## 1 | 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, 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 behaviour [4, 6, 7]. For instance, in payment systems, to ensure strong customer authentication, it is necessary to implement adequate security features<sup>1</sup> 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 unauthorised parties discovering, disclosing, and using elements such as algorithm specifications, key length, and information entropy [9]. When Multi-Factor Authentication is requested, using Seamless biometrics, as behavioural, improves

<sup>1</sup>[http://data.europa.eu/eli/reg\\_del/2018/389/oj](http://data.europa.eu/eli/reg_del/2018/389/oj).

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *IET Biometrics* published by John Wiley & Sons Ltd.

the security without decreasing the User Experience (UX). Increasing the 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.

Behavioural biometrics involves measuring a user's behavioural tendencies, which can include gait, human activity, voice recognition, signature verification, keystroke dynamics, mouse dynamics, and Graphical User Interface usage analysis [12]. According to Bailey et al., behavioural 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 analysing 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 UX. 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 behavioural modality.

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

The contributions of the proposed paper are numerous. The proposed research uses an image-based architecture (for a chosen behavioural biometric modality: gait analysis). A deep learning process for user authentication based on human activity is proposed. We consider only one behavioural 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 organised as follows. Section 2 contains related works on authentication systems from human activity. Section 3 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 4 draws the experimental protocol. Section 5 details the experiments on benchmark datasets and the results we obtained. Section 6 gives the conclusions of this work and some perspectives.

## 2 | 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 nonintrusive authentication employ user behavioural features such as gait, touch screen operations, hand gestures, keyboard patterns, speech, or signature movements to generate behavioural 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 Equal Error Rate (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 [40], Bayesian Net classifiers [41], and Naïve Bayes [42].

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

Table 1 lists the aims of human activity. Human activity can be used for different goals (identification, authentication, and 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 1** 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*

\*That the manuscript is positioned on the Biometric verification/authentication of subjects in the context of human activity objectives.

Table 2 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 2 focusses on activity recognition, and less work focusses 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.

### 3 | 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 Ref. [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.

#### 3.1 | 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 study, we use a function that helps us transform behavioural biometric signals, which can be considered as time series, into an image, that is, 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* 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.

For instance, consider the University of California, Irvine Human Activity Recognition (UCI-HAR) dataset<sup>3</sup>, 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  $34 \times 34$ , where  $n = 34$ . The matrix is displayed with false colours representing distance values.

We finally have a 2D image in Red, Green, and Blue (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 colour image). The transformation is performed on each attempt of each user for each human activity.

#### 3.2 | 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<sup>4</sup>, 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 a result of a binary classification problem (genuine or impostor).

Table 3 summarises the architecture and the optimisation 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.

#### 3.3 | 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

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

- The Euclidean distance

$$d^2 = (x_s - x_t)(x_s - x_t)' \quad (2)$$

- The Cosine distance

$$d = 1 - \frac{x_s x'_t}{\sqrt{(x_s x'_s)(x_t x'_t)}} \quad (3)$$

<sup>2</sup> <https://fr.mathworks.com/help/stats/squareform.html>.

<sup>3</sup> <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>.

<sup>4</sup> <https://image-net.org>.

**TABLE 2** 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 neighbour	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 behaviour	Ambient sensors	98.0%	-
[25]	Active learning to recognise human activity using Smartwatch	RF, extra trees, Naive Bayes, Logistic Regression, SVM	Running, walking, standing, sitting, lying down	Smartwatch	93.3%	-
[26]	Recognising 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%]

(Continues)

TABLE 2 (Continued)

Paper	Approach	Method	Activity	Input source	Accuracy	EER
[35]	User verification	Manhattan distance	Hand movement	Keyboard	[89.00%–94.00%]	[06.00%–11.00%]

Abbreviations: AIBO, Artificial Intelligence Robot and Companion; ANN, Artificial Neural Network; CNN, Convolution Neural Network; DTW, Dynamic Time Warping; EER, Equal Error Rate; GMM, Gaussian Mixture Model; GUI, Graphical User Interface; HMM, Hidden Markov Model; KF, Kalman Filter; RF, Random Forest; RNN, Recurrent Neural Network; SVC, Support Vector Classifier; SVM, Support Vector Machine.

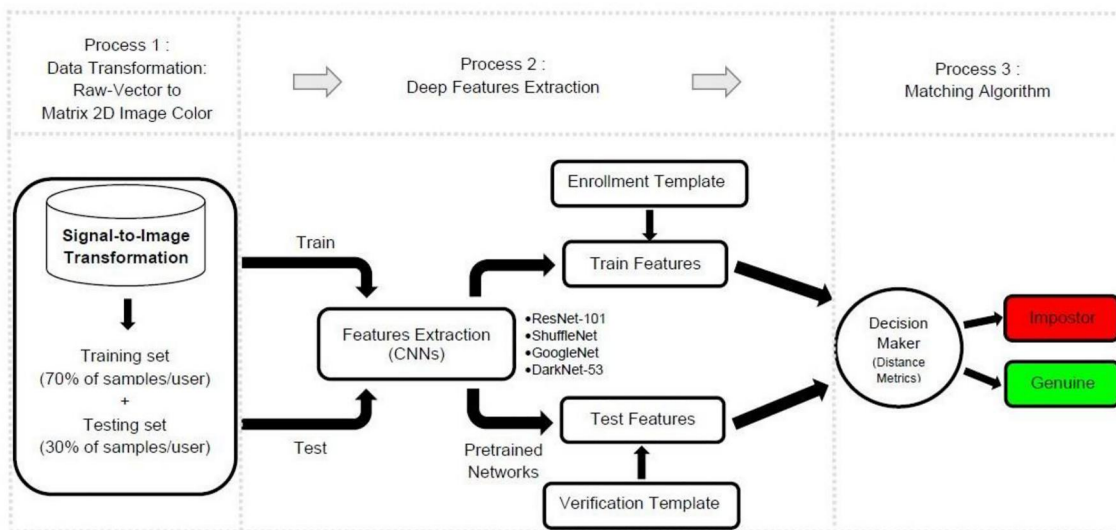


FIGURE 1 Architecture of the authentication system.

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
Initialise counters cptra = 1 cpter = 1
Compute intraclass and interclass scores
i = 1; j = 1; i = 1; j = 1; i = 1; j = 1
INTRA(cptra) = pdist([data(M*(i-1)+1, :);
data(M*(i-1)+j, :)], distance)
cptra = cptra + 1
INTER(cpter) = pdist([data(M*(i-1)+1, :);
data(M*(i-1)+j, :)], distance)
cpter = cpter + 1
```

**Return** INTRA, INTER

Algorithm 1 describes the procedure for computing INTRA scores for similarity and INTER scores for dissimilarity. These scores are used to compute the False Match Rate (FMR), False Non-Match Rate (FNMR), and EER scores later on.

## 4 | 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.

### 4.1 | 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 six 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 by using the accelerometer and gyroscope (3-axial raw signals with *tAcc-XYZ* and *tGyro-XYZ*) sensors at a frequency of 50 Hz. 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 labelled. UCI-HAR contains 10,299 samples.

Table 4 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 3) is applied to obtain a 2D colour image. To the best of our knowledge, such transformation with the *squareform()* function applied to the

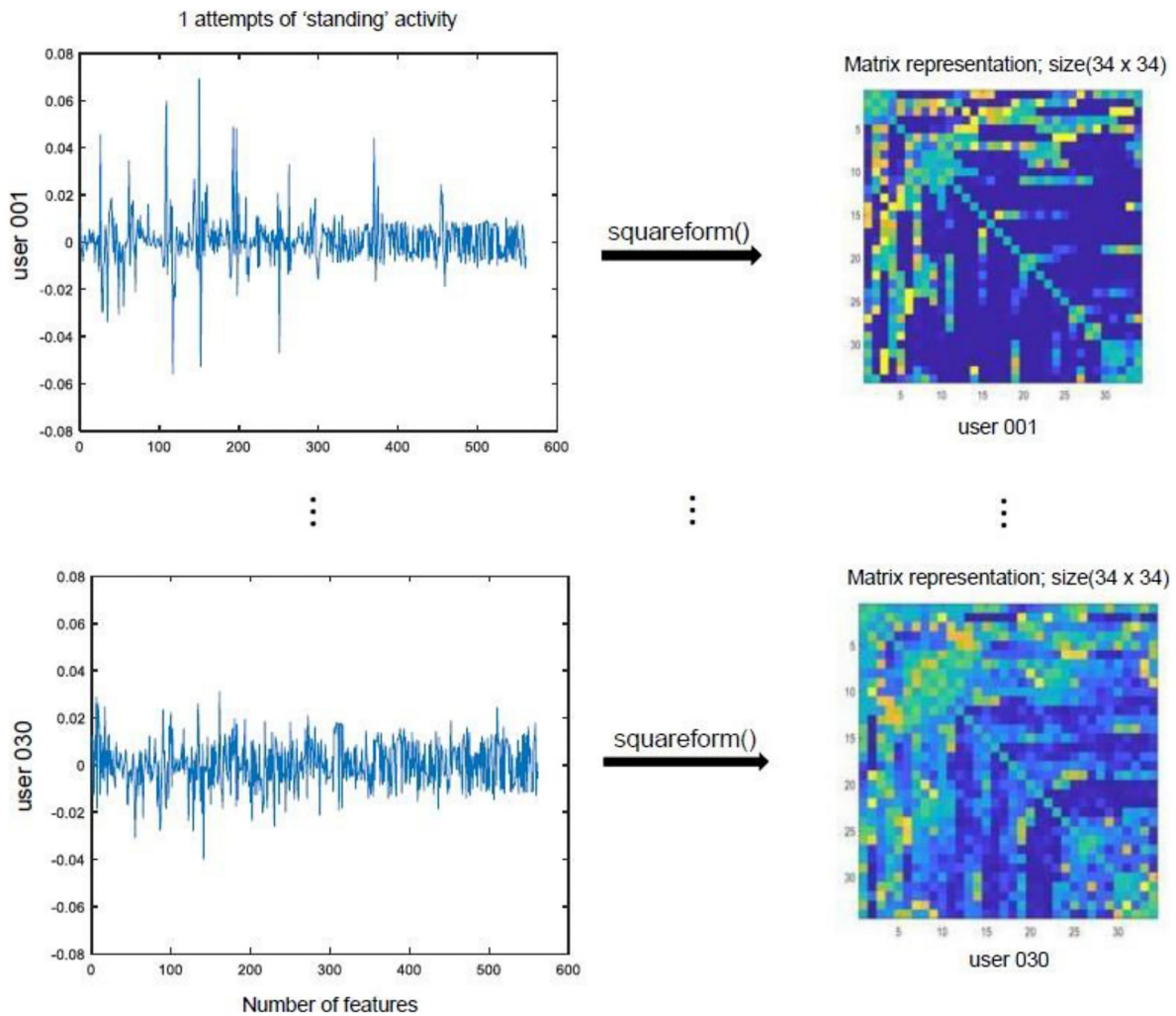


FIGURE 2 Examples of the obtained results when the signal-to-image transformation is applied.

TABLE 3 Architectures and optimisations hyper-parameters for the deep learning approaches.

Models	#Layers	Depth	Image input size	Activate	Normalise	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

Abbreviation: SGDM, Stochastic Gradient Descent with Momentum.

TABLE 4 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

Abbreviation: UCI-HAR, University of Californy, Irvine Human Activity Recognition.

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.

## 4.2 | 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 Ref. [3].

## 4.3 | 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 3.3 are applied to evaluate the degree of similarity between the activities of each user.

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

- 1) The FMR is the proportion of a specified set of completed nonmated 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 *nonmatch*.

The 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 minimise 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 have been realised on an Intel Core i5-9600K Central Processing Unit 3.70 GHz computer equipped with 16.00 GB of RAM with MatLab.

## 5 | RESULTS AND DISCUSSION

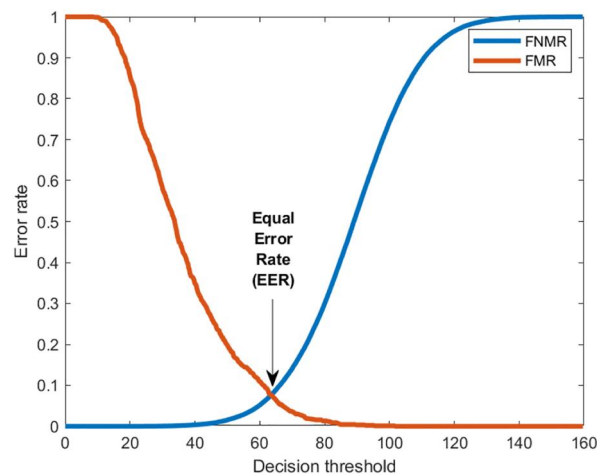
In this section, we present the experimental results we obtained. We tried to structure them by addressing some questions concerning the performance of the proposed method on

such behavioural biometric datasets. Note that we considered 70% of user samples (attempts per subject) for the learning phase and 30% for the testing one.

### 5.1 | Which performance can we expect on a larger dataset ?

First, we consider all the six activity sub-datasets defined as the UCI-HAR dataset (**Fusion of features**). Table 5 draws the obtained results for the user verification task, considering the three distances.

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 the EER value.



**FIGURE 3** Relationship between FMR, FNMR and EER. EER, Equal Error Rate; FMR, False Match Rate; FNMR, False Non-Match Rate. Source: [3].

**TABLE 5** EER value on the HAR dataset for the three tested distances.

Models	EER <sub>manathan</sub> (%)	EER <sub>euclidean</sub> (%)	EER <sub>cosine</sub> (%)
ResNet-101	22.69	17.71	<b>12.48</b>
ShuffleNet	14.77	14.63	<b>11.57</b>
GoogleNet	14.88	14.56	<b>13.52</b>
DarkNet-53	17.46	14.31	<b>11.72</b>

Note: The bold values in Table 5 show that, whatever the deep architecture, the distance that minimizes the EER value is the cosine distance. This allows us to consider only the cosine distance in the remainder of this work.

Abbreviation: EER, Equal Error Rate.



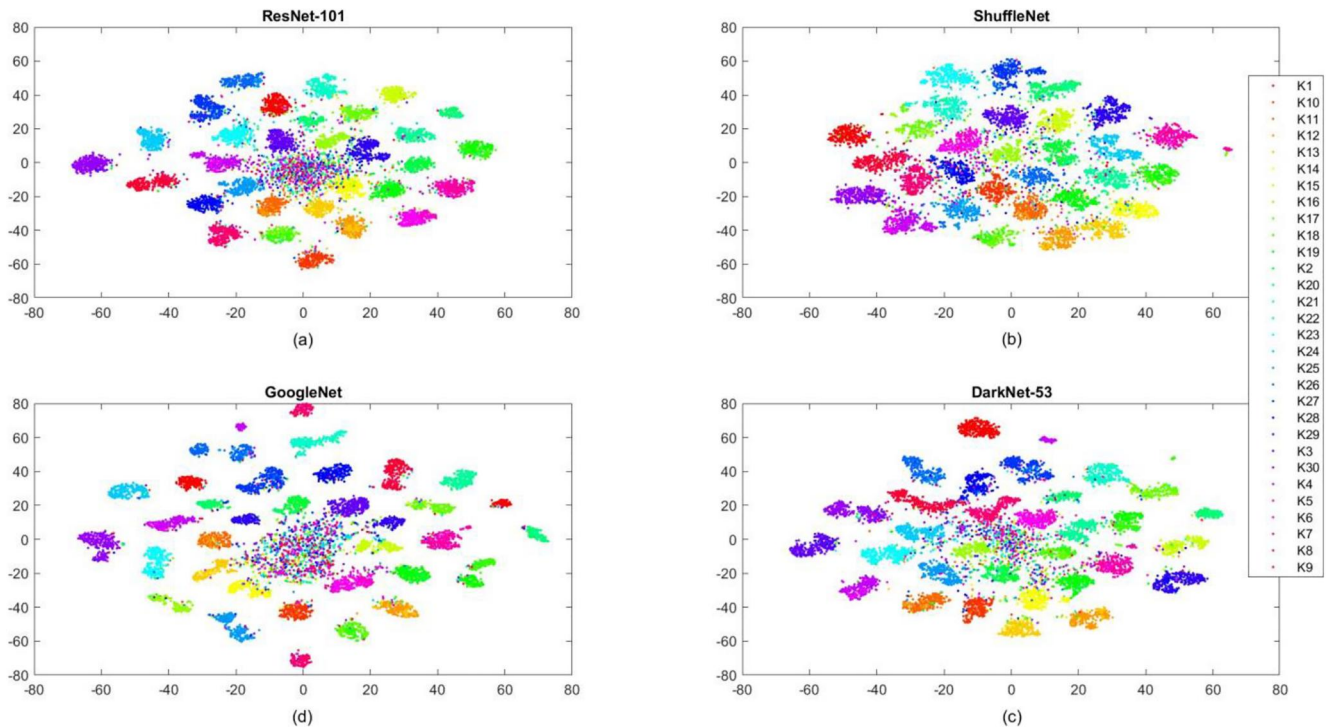


FIGURE 4 Visual inspection of deep features projection from (a) ResNet-101, (b) ShuffleNet, (c) DarkNet-53, and (d) GoogleNet.

## 5.2 | How well can we perform on each activity separately?

In this section, we consider each activity separately as shown in Table 4 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), and walking upstairs (1944 samples for the 30 subjects). We illustrate the four architectures (namely ResNet-101, ShuffleNet, GoogleNet, and DarkNet-53), and we draw the model on each sub-datasets separately.

This is illustrated by block 1 in Figure 5. The best model among the four deep networks for each activity in terms of the 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.

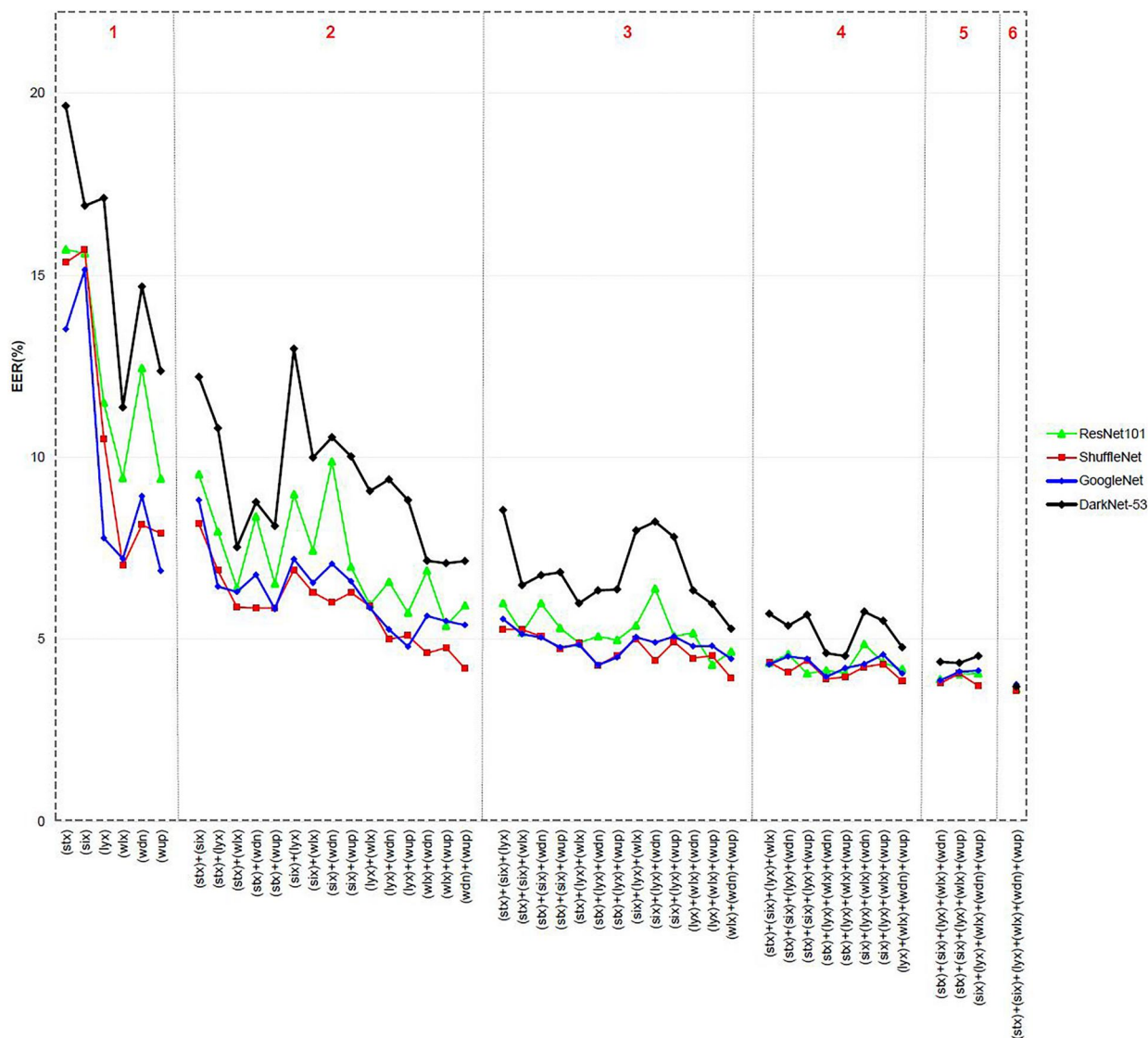
We note that we do not have the same performance from one activity to another. So, using 70% of samples (attempts per subject) for the generation of the reference template does not provide exceptional results (with an EER value between 06.88% and 19.65%) as shown in block 1 in Figure 5. Obviously, if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

## 5.3 | 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 6 shows the obtained results if we used all the six activities (i.e. simulating a user achieving six activities to be authenticated). ShuffleNet comes out as the best method with an EER score of 03.58% as presented in Table 6. ShuffleNet is ahead of ResNet-101 (03.63%), GoogleNet (03.76%) and DarkNet-53 (03.70%).

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 two activities, we obtain an EER value between [04.20% – 12.98%] illustrated by block 2 in Figure 5.
- If we have three activities, we have an EER value between [03.94% – 08.55%] represented in block 3.
- If we use four activities, we have an EER value around [03.85% – 05.71%] depicted by block 4.
- If we use five activities, we have an EER value around [03.72% – 04.54%] shown by block 5.
- If we use six activities (*laying + sitting + standing + walking + walking downstairs + walking upstairs*), we get an EER value around [03.58% – 03.76%] depicted by block



**FIGURE 5** EER rate on deep architectures for the multi-instance biometric system. In block 1, we have  $\{(stx), (six), (lyx), (wlx), (wdn), (wup)\}$  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,  $\{(stx)+(six)+(lyx)+(wlx)+(wdn)+(wup)\}$ . EER, Equal Error Rate.

**TABLE 6** Performance evaluation on the multi-instance biometric system by fusion of features and scores level on the UCI-HAR dataset.

Models (EER <sub>cosine</sub> )	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

Abbreviation: UCI-HAR, University of Californy, Irvine Human Activity Recognition.

5. This shows that 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, and 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].

## 5.4 | 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 unauthorised access [1].

The purpose of this work is to analyse several information from user activity in order to authenticate himself/herself. A

comparative analysis of the four architectures on the UCI-HAR dataset allows us to identify the best performance for a continuous authentication. From Table 5, it is provided by the ShuffleNet architecture with an EER value equal to 11.57%. Figure 6 shows a visual inspection of features (raw vs. 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. Table 6 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.

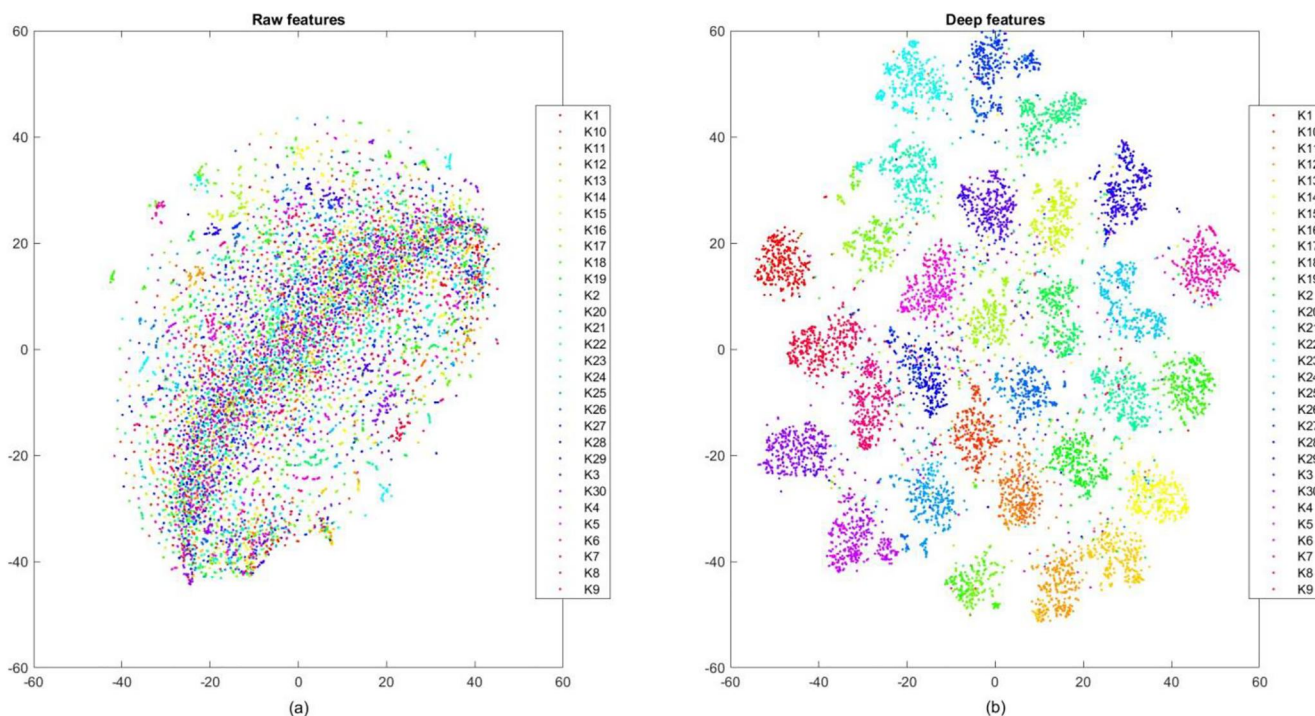
In the literature, several works notably [51–58] shown in Table 2 and others in Table 7 have been carried out only on the recognition of activities (where the target is standing, sitting, laying, walking, walking downstairs, and walking upstairs) from the UCI-HAR dataset. The used methods are respectively Deep CNN-LSTM with Self-Attention (accuracy = 93.11%), linear Support Vector Classifier (accuracy = 96.50%), LSTM-CNN (accuracy = 95.78%), Support Vector Machine (accuracy = 97.12%), Lego-CNN (accuracy = 96.90%), LSTM (accuracy = 97.40%), and Convolution Neural Network (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 focusses on activity-based user verification.

We can compare our results with research works that has been performed on the UCI-HAR dataset in Table 8. 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 Deep-ConvLSTM 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.

## 6 | CONCLUSION AND PERSPECTIVES

This survey is based on a new method for user authentication by analysing human activities in this study. 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 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 EER for continuous smartphone authentication utilising



**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.

**TABLE 7** 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%]	-

Abbreviations: CMU, Cumulative Match Curve; CNN, Convolution Neural Network; EER, Equal Error Rate; SVM, Support Vector Machine; UCF, University of Central Florida; UCI-HAR, University of California, Irvine Human Activity Recognition.

**TABLE 8** 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	Mantjarvi et al. [18]	2005	Correlation coefficients	-	7
Biometric gait data	Muaazz et al. [20]	2013	SVM	-	33.30
Mobile gait data	Zhong et al. [21]	2015	Nearest neighbour	-	07.22

Abbreviations: EER, Equal Error Rate; RNN, Recurrent Neural Network; SVM, Support Vector Machine; UCI-HAR, University of California, Irvine Human Activity Recognition; WISDM, Wireless Sensor Data Mining.

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, and 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.

## AUTHOR CONTRIBUTIONS

**Yris Brice Wandji Piugie:** Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Resources; Software; Validation; Visualisation; Writing – original draft; Writing – review & editing. **Christophe Charrier:** Conceptualisation; Formal analysis; Funding acquisition; Project administration; Resources; Supervision; Validation; Writing – original draft. **Joël Di Manno:** Formal analysis; Funding acquisition;

Investigation; Project administration; Resources; Supervision; Validation; Writing – original draft. **Christophe Rosenberger:** Conceptualisation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing – original draft.

## ACKNOWLEDGEMENTS

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.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in UCI Machine Learning Repository at reference number <https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones>.

## ORCID

Yris Brice Wandji Piugie  <https://orcid.org/0000-0002-6258-6073>

## REFERENCES

- Mekruksavanich, S., Jitpattanukul, A.: Deep learning approaches for continuous authentication based on activity patterns using mobile sensing. *Sensors* 21(22), 7519 (2021). <https://doi.org/10.3390/s21227519>
- Nugier, C., et al.: An upcycling tokenization method for credit card numbers. In: *SECRYPT 2021-18th International Conference on Security and Cryptography* (2021)
- Piugie, Y.B.W., et al.: Keystroke dynamics based user authentication using deep learning neural networks. In: *2022 International Conference on Cyberworlds (CW 2022)* (2022)
- Piugie, Y.B.W., et al.: How artificial intelligence can be used for behavioral identification? In: *2021 International Conference on Cyberworlds (CW)* (2021)
- Visvizi, A., et al.: Tweeting and mining oecd-related microcontent in the post-truth era: a cloud-based app. *Comput. Hum. Behav.* 107, 105958 (2020). <https://doi.org/10.1016/j.chb.2019.03.022>
- Rasekh, A., Chen, C.-A., Lu, Y.: Human activity recognition using smartphone. *arXiv preprint arXiv:1401.8212* (2014)
- Piugie, Y.B.W., et al.: Denoising of electroencephalographic signals by canonical correlation analysis and by second-order blind source separation. In: *2019 IEEE Africon*, pp. 1–8. IEEE (2019)
- Cherrier, E.: Authentification biométrique: comment (ré) concilier sécurité, utilisabilité et respect de la vie privée? Ph.D. dissertation. Normandie Université (2021)
- Migdal, D.: Contributions to Keystroke Dynamics for Privacy and Security on the Internet. Ph.D. dissertation. Normandie Université (2019)
- Yohan, M., Hanry, H., Dion, D.: Keystroke dynamic classification using machine learning for password authorization. In: *3rd International Conference on Computer Science and Computational Intelligence, Procedia Computer Science* (2018)
- Kim, D.I., Lee, S., Shin, J.S.: A new feature scoring method in keystroke dynamics-based user authentications. *IEEE Access* 8, 27901–27914 (2020). <https://doi.org/10.1109/access.2020.2968918>
- Bailey, K.O., Okolica, J.S., Peterson, G.L.: User identification and authentication using multi-modal behavioral biometrics. *Comput. Secur.* 43, 77–89 (2014). <https://doi.org/10.1016/j.cose.2014.03.005>
- Marsico, M.D., Mecca, A.: Biometric walk recognizer: gait recognition by a single smartphone accelerometer. *Multimed. Tool. Appl.* 76(4), 4713–4745 (2017). <https://doi.org/10.1007/s11042-016-3654-1>
- Barra, P., et al.: 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*, pp. 180–190. Springer (2019)
- Patel, V.M., et al.: Continuous user authentication on mobile devices: recent progress and remaining challenges. *IEEE Signal Process. Mag.* 33(4), 49–61 (2016). <https://doi.org/10.1109/msp.2016.2555335>
- Zhang, M.: Gait activity authentication using LSTM neural networks with smartphone sensors. In: *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pp. 456–461. IEEE (2019)
- Gao, G., et al.: Hierarchical deep CNN feature set-based representation learning for robust cross-resolution face recognition. In: *IEEE Transactions on Circuits and Systems for Video Technology* (2020)
- Mantjarvi, J., et al.: 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*, pp. ii–973. IEEE (2005)
- Thang, H.M., et al.: Gait identification using accelerometer on mobile phone. In: *2012 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pp. 344–348. IEEE (2012)
- Muaz, M., Mayrhofer, R.: An analysis of different approaches to gait recognition using cell phone based accelerometers. In: *Proceedings of International Conference on Advances in Mobile Computing & Multimedia*, pp. 293–300 (2013)
- Zhong, Y., Deng, Y., Meltzner, G.: Pace independent mobile gait biometrics. In: *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–8. IEEE (2015)
- Alex, P.M.D., et al.: Research on human activity identification based on image processing and artificial intelligence. *Int. J. Eng. Technol.* 7(3.27), 174 (2018). <https://doi.org/10.14419/ijet.v7i3.27.17754>
- Jaouedi, N., Boujnah, N., Bouhleb, M.S.: A new hybrid deep learning model for human action recognition. *J. King Saud Univ. Comput. Inf. Sci.* 32(4), 447–453 (2020). <https://doi.org/10.1016/j.jksuci.2019.09.004>
- Antón, M.Á., et al.: Non-invasive ambient intelligence in real life: dealing with noisy patterns to help older people. *Sensors* 19(14), 3113 (2019). <https://doi.org/10.3390/s19143113>
- Shahmohammadi, F., et al.: Smartwatch based activity recognition using active learning. In: *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 321–329. IEEE (2017)
- Anguita, D., et al.: 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*, pp. 437–442 (2013)
- Sanchez Guinea, A., Sarabchian, M., Mühlhäuser, M.: Improving wearable-based activity recognition using image representations. *Sensors* 22(5), 1840 (2022). <https://doi.org/10.3390/s22051840>
- Sarkar, A., Hossain, S., Sarkar, R.: Human activity recognition from sensor data using spatial attention-aided CNN with genetic algorithm. *Neural Comput. Appl.* 35(7), 1–27 (2022). <https://doi.org/10.1007/s00521-022-07911-0>
- Junejo, I.N., et al.: Cross-view action recognition from temporal self-similarities. In: *European Conference on Computer Vision*, pp. 293–306. Springer (2008)
- Zhong, Y., Deng, Y.: Sensor orientation invariant mobile gait biometrics. In: *IEEE International Joint Conference on Biometrics*, pp. 1–8. IEEE (2014)
- Körner, M., Denzler, J.: Temporal self-similarity for appearance-based action recognition in multi-view setups. In: *International Conference on Computer Analysis of Images and Patterns*, pp. 163–171. Springer (2013)
- Sun, C., et al.: Exploring sparseness and self-similarity for action recognition. *IEEE Trans. Image Process.* 24(8), 2488–2501 (2015). <https://doi.org/10.1109/tip.2015.2424316>
- Zareen, F.J., Jabin, S.: Authentic mobile-biometric signature verification system. *IET Biom.* 5(1), 13–19 (2016). <https://doi.org/10.1049/iet-bmt.2015.0017>
- Gafurov, D., Helkala, K., Söndrol, T.: Biometric gait authentication using accelerometer sensor. *J. Comput.* 1(7), 51–59 (2006). <https://doi.org/10.4304/jcp.1.7.51-59>
- Parkinson, S., et al.: Password policy characteristics and keystroke biometric authentication. *IET Biom.* 10(2), 163–178 (2021). <https://doi.org/10.1049/bme2.12017>
- Sitová, Z., et al.: HMOG: new behavioral biometric features for continuous authentication of smartphone users. *IEEE Trans. Inf. Forensics Secur.* 11(5), 877–892 (2015). <https://doi.org/10.1109/tifs.2015.2506542>
- Giorgi, G., Saracino, A., Martinelli, F.: Using recurrent neural networks for continuous authentication through gait analysis. *Pattern Recogn. Lett.* 147, 157–163 (2021). <https://doi.org/10.1016/j.patrec.2021.03.010>
- Zheng, N., et al.: You are how you touch: user verification on smartphones via tapping behaviors. In: *2014 IEEE 22nd International Conference on Network Protocols*, pp. 221–232. IEEE (2014)
- Trojahn, M., Ortmeier, F.: Toward mobile authentication with keystroke dynamics on mobile phones and tablets. In: *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, pp. 697–702. IEEE (2013)
- Pal, S.K., Mitra, S.: *Multilayer Perceptron, Fuzzy Sets, Classification* (1992)

41. Kohavi, R., et al.: Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid. In: *Kdd*, Vol. 96, pp. 202–207 (1996)
42. Neverova, N., et al.: Learning human identity from motion patterns. *IEEE Access* 4, 1810–1820 (2016). <https://doi.org/10.1109/access.2016.2557846>
43. De Marsico, M., De Pasquale, D., Mecca, A.: Embedded accelerometer signal normalization for cross-device gait recognition. In: 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–5. IEEE (2016)
44. Chen, Z., et al.: 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), pp. 649–654. IEEE (2021)
45. Migdal, D.: Contributions to Keystroke Dynamics for Privacy and Security on the Internet. Theses. Normandie Université (2019). [Online]. <https://tel.archives-ouvertes.fr/tel-02518436>
46. Vasconcelos, C., et al.: Impact of aliasing on generalization in deep convolutional networks. *arXiv preprint arXiv:2108.03489* (2021)
47. Maheshwary, S., Ganguly, S., Pudi, V.: 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. Aversano, L., et al.: Continuous authentication using deep neural networks ensemble on keystroke dynamics. *PeerJ Comput. Sci.* 7, e525 (2021). <https://doi.org/10.7717/peerj-cs.525>
49. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115(3), 211–252 (2015). <https://doi.org/10.1007/s11263-015-0816-y>
50. ISO: Information technology — biometric performance testing and reporting — part 1: principles and framework. In: *International Organization for Standardization*, Geneva, CH, Standard ISO/IEC 19795-1:2021 (2021)
51. Khatun, M.A., et al.: Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. *IEEE J. Trans. Eng. Health Med.* 10, 1–16 (2022). <https://doi.org/10.1109/jtehm.2022.3177710>
52. Khare, S., Sarkar, S., Totaro, M.: Comparison of sensor-based datasets for human activity recognition in wearable iot. In: 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), pp. 1–6. IEEE (2020)
53. Mahmud, T., et al.: A novel multi-stage training approach for human activity recognition from multimodal wearable sensor data using deep neural network. *IEEE Sensor. J.* 21(2), 1715–1726 (2020). <https://doi.org/10.1109/jsen.2020.3015781>
54. Xia, K., Huang, J., Wang, H.: LSTM-CNN architecture for human activity recognition. *IEEE Access* 8, 56855–56866 (2020). <https://doi.org/10.1109/access.2020.2982225>
55. Jain, A., Kanhangad, V.: Human activity classification in smartphones using accelerometer and gyroscope sensors. *IEEE Sensor. J.* 18(3), 1169–1177 (2017). <https://doi.org/10.1109/jsen.2017.2782492>
56. Tang, Y., et al.: Layer-wise training convolutional neural networks with smaller filters for human activity recognition using wearable sensors. *IEEE Sensor. J.* 21(1), 581–592 (2020). <https://doi.org/10.1109/jsen.2020.3015521>
57. Tufek, N., et al.: Human action recognition using deep learning methods on limited sensory data. *IEEE Sensor. J.* 20(6), 3101–3112 (2019). <https://doi.org/10.1109/jsen.2019.2956901>
58. Zebin, T., et al.: Design and implementation of a convolutional neural network on an edge computing smartphone for human activity recognition. *IEEE Access* 7, 133509–133520 (2019). <https://doi.org/10.1109/access.2019.2941836>

**How to cite this article:** Wandji Piugie, Y.B., et al.: Deep features fusion for user authentication based on human activity. *IET Biome.* 12(4), 222–234 (2023). <https://doi.org/10.1049/bme2.12115>