Mobile Passive Authentication through Touchscreen and Background Sensor Data

Giuseppe Stragapede
Biometrics and Data Pattern Analytics
Universidad Autonoma de Madrid
Madrid, Spain
giuseppe.stragapede@uam.es

Ruben Vera-Rodriguez
Biometrics and Data Pattern Analytics
Universidad Autonoma de Madrid
Madrid, Spain
ruben.vera@uam.es

Ruben Tolosana
Biometrics and Data Pattern Analytics
Universidad Autonoma de Madrid
Madrid, Spain
ruben.tolosana@uam.es

Aythami Morales
Biometrics and Data Pattern Analytics
Universidad Autonoma de Madrid
Madrid, Spain
aythami.morales@uam.es

Alejandro Acien
Biometrics and Data Pattern Analytics
Universidad Autonoma de Madrid
Madrid, Spain
alejandro.acien@uam.es

Gael Le Lan
Orange Labs
Cesson-Sevigne, France
gael.lelan@orange.com

Abstract—The security and usability shortcomings of current mobile user authentication systems based on PIN codes, fingerprint, and face recognition are well known. To overcome such limitations, the present work focuses on the comparative analysis of unimodal and multimodal behavioral biometric traits suitable for mobile passive authentication, such as touchscreen data during separate gestures (keystroke, scrolling, drawing a number, tapping on the screen), and background sensor data (accelerometer, gravity sensor, gyroscope, linear accelerometer, magnetometer).

This paper carries out a performance evaluation over one of the most complete and challenging databases to date with mobile user interaction data, HuMIdb, with 600 subjects. For each individual modality, we propose a separate RNN (Recurrent Neural Network) trained with semi-hard triplet loss. In addition, we perform the fusion of the different modalities at score level. Our results show that the best performing tasks are keystroke and drawing a number, whereas the most discriminative background sensor is the magnetometer. Additionally, the fusion of modalities is very beneficial, consistently reducing the Equal Error Rates (EER) by half (ranging from 5% to 13% depending on the modality combination).

Index Terms—behavioral biometrics, passive authentication, mobile devices, human computer interaction

I. INTRODUCTION

The amount of operations carried out in the digital domain is increasing, especially in the area of identity and access management. As we are entrusting our devices with such a great amount of personal information, a further need for privacy and security arises, compared to traditional systems.

As of today, most mobile authentication methods rely on knowledge of *secrets*, such as passwords, PIN codes or graphic patterns (according to the *what-you-know* authentication paradigm). It is well known that such methods are vulnerable to *shoulder-surfing*, *guess* and *smudge* attacks [1], [2]. Moreover, it has been shown that the average subject spends about 2.9% of their total usage time for knowledge-based authentication [3].

A user's biometric trait can be used to perform personal authentication (according to the *what-you-are* paradigm). The first step of biometric authentication systems consists in the user enrollment: the user's personal traits are captured, transformed to fit a predetermined template after the extraction of meaningful features, aiming to maximize users' distinctiveness. Then, upon a new access request, the system will perform a verification task to determine whether the current user's information matches the stored information. The main advantages of biometric authentication consist in the fact that it requires no mnemonic effort, additional devices nor are they vulnerable to the above-mentioned attacks. In light of this, current mobile authentication systems rely on also on physiological biometrics[1], as in the case of fingerprint or face recognition systems. Such systems, however, are prone to presentation attacks (spoofing), although they do not require mnemonic efforts [4], [5]. Furthermore, they all typically do not provide prolonged protection over the entire device usage. The user would have, in fact, to frequently interrupt their activity to have their fingerprint scanned or to type in their pass code, whereas repeated face verification appears impractical due to hardware constraints. In other words, if an attacker gains access to the device, they can stay authenticated as long as the device remains active [6].

Consequently, it is necessary to develop more secure and usable methods to verify the identity match of the device owner and legitimate user. In this scenario, Continuous Au-

[1]Physiological biometrics refer to the biological characteristics that allow to identify an individual.

thentication[2] (CA) approaches are designed to overcome these limitations by constantly verifying the user's biometric traits in a *passive* way, i.e. without requiring them to complete any predetermined task [7], [8]. An opportunity for CA on mobile devices is offered by behavioral biometric traits. Mobile devices are in fact equipped with several sensors, capable of acquiring a vast amount of biometric information [9]. As opposed to physiological biometrics, they are a broad category which includes all the means that allow or help in discriminating among individuals based on the *way* activities are performed, such as gait, typing, scrolling, signature, etc. [10]–[16]. They provide stronger security guarantees compared to physical traits, as they demand advanced technical skills to be spoofed [17]. Moreover, combining different modalities to achieve a *multimodal*[3] biometric system has proven to be beneficial in terms of robustness, immunity to noise, universality, and security, at a cost of increased complexity [7], [18], [19]. Dealing with data acquired by numerous sensors and sources for passive authentication is a complex process that entails multiple data pre-processing methods, development of models, and verification techniques.

Several studies in the literature have focused on achieving continuous user authentication throughout such *modalities* [20], [21]. Among these, Deb *et al.* proposed a contrastive loss-based Siamese RNN (Recurrent Neural Network) architecture with LSTM (Long Short-Term Memory) units for passive authentication [22]. On a small self-collected dataset, they fused eight different modalities, achieving 96.47% of True Acceptance Rate (TAR) at a False Acceptance Rate (FAR) of 0.1% within 3 seconds. Based on a similar network architecture, Acien *et al.* first explored the touch data using the same database considered for the current study, the HuMIdb (Sec. II), reaching accuracies up to 87% for user authentication based on a simple and fast touch gesture [23].

In the current work, we perform a comprehensive analysis of individual behavioral biometric traits suitable for the application of CA through the interaction with mobile devices. In comparison with previous works [20]–[23], dedicated user-device interaction tasks are considered (keystroke, scroll up, scroll down, drawing a number with the finger, tap on the screen). In addition, we explore the complementarity between task-dependent features and background sensors features (accelerometer, gravity sensor, gyroscope, linear accelerometer, magnetometer). For each individual modality, we implement a separate LSTM RNN trained with semi-hard triplet loss [24], comparing the biometric performance of individual modalities and also their fusion at score level. To this end, we carry out a first benchmark of the HuMIdb (Human Mobile Interaction database), a novel database that comprises more than 5GB from a wide range of mobile sensor data acquired under unsupervised scenario [25]. Finally, we assess of the impact of the quantity of enrollment data on the recognition performance.

---

[2]The terms *continuous* authentication, *implicit* authentication and *transparent* authentication have been used interchangeably in the literature [7].

[3]In the context of user authentication, a source of biometric information useful to recognize an individual is commonly referred to as *modality*.

II. HuMIdb Database

The Human Machine Interaction database (HuMIdb[4]) is a publicly available database which includes 14 sensors during natural human-mobile interaction performed by 600 subjects [25]. The acquisition was carried out over five separate sessions with at least a 24-hour gap among them in order to account for intra-subject variability. An Android application was implemented to acquire the signals while the subjects completed eight tasks with their own smartphones in an unsupervised scenario. The participants were recruited from 14 countries (52.2% European, 47.0% American, 0.8% Asian) using 179 different device models in the attempt to obtain more diverse data than previous state-of-the-art databases. The tasks include keystroke, scroll up and scroll down gestures, tap gestures, draw in the air with the smartphone a circle and a cross, say "I am not a robot", and draw with the finger the digits 0 to 9 over the touchscreen. Additionally, all tasks have a right swipe button that is acquired in addition to the swipe patterns. The tasks considered within this study, as named within the database sub-directories, are "KEYSTROKE", "SCROLLUP", "SCROLLDOWN", "FINGER_8", "TOUCH". The keystroke data were acquired in a fixed-text scenario corresponding to the sentence "En un lugar de la Mancha, de cuyo nombre no quiero acordarme". All mistakes made by the subject while typing were acquired. The scroll up and the scroll down gestures were acquired in portrait mode. The number "8" (as the other numbers) was drawn on the screen by subjects with their finger. Finally, for the tapping task, the subjects had to tap on a series of buttons in predetermined locations of the screen.

III. Pre-Processing and Feature Extraction

Before the training, validation, and final testing of the the models, the raw data contained in the database is pre-processed according to the following description.

A. Background Sensor Data

Down-sampling is applied to the different signals per axis as this has proven to be beneficial to the user authentication performance (the down-sampling ratio used is $D = 16$). In this operation, we consider the average value of D consecutive samples.

By extracting additional features from the raw data (x, y, z), such as the first- and second-order derivatives and Fast Fourier Transform (FFT), features pertaining to a single timestamp are arranged into 12-dimensional vector (4-dimensional in the case of the gravity sensor, as in this case the sensor output is one-dimensional):

$$[x, y, z, x', y', z', x'', y'', z'', \mathit{fft}(x), \mathit{fft}(y), \mathit{fft}(z)]$$

They are grouped in *M-sample arrays* of shape $M \times 12$ that represent a single time window (in the case of background sensors, $M = 150$). For each one of the five acquisition sessions, one time window comprising M samples was obtained.

---

[4]https://github.com/BiDAlab/HuMIdb

Then, in each time window, data normalization is carried out by subtracting the mean and dividing by the standard deviation of the time domain signals along each axis. In order to compute the FFT, the raw time-domain signals for each axis are arranged in fixed-size time windows of 2,400 samples. This value corresponds to the multiplication of the array size M by the down-sampling ratio D. Then, the normalization step is performed, followed by the computation of the FFT. An additional normalization step follows, as the lower frequency values have typically higher values in terms of magnitude than the ones in the rest of the spectrum. Then, the down-sampling is performed to finally obtain evenly-distributed spectral information arranged in the same shape as the time-domain data.

B. Touch Data

The tasks considered are scroll up, scroll down, drawing the number "8" with the finger and tapping on the screen. The number 8 was selected as it is one of the most complex graphic patterns among numbers. The raw data consist in x and y coordinates of the touch, and the pressure p applied. The pre-processing pipeline is similar to the case of background sensor signals, including the first- and second-order derivatives and the FFT, but without down-sampling. The x and y data are divided by the height and width values of the screen of the particular device, and the pressure p is normalized by subtracting the mean and dividing by the standard deviation of the time domain signals along each axis. For the scrolling tasks, the sequence length is set to $M = 100$, whereas for tapping $M = 15$, as such gestures are generally shorter and the number of samples acquired is lower.

C. Keystroke Data

The only available data in the HuMIdb corresponds to the timestamp of the press gesture and the key pressed. Therefore, the features considered are the inter-press time and the normalized value of the ASCII code.

[*inter-press time, key pressed*]

The data were again arranged into fixed-size time windows of $M = 150$ samples, obtaining one time window per acquisition session.

IV. System Description

A. Learning Architecture

The authentication system is based on a LSTM RNN, a deep learning architecture particularly apt to capture long term dependencies in time domain sequences [26]. The network considered comprises of two layers of 64 units with a *tanh* activation function. Between the LSTM layers, batch normalization and a dropout rate of 0.5 are implemented to avoid overfitting. Additionally, each LSTM layer has a recurrent dropout rate of 0.2.

The output of the models for a given M-sample time window consists of a feature embedding, in other words, an E-dimensional array of real values ($E = 64$). The scores are calculated in terms of Euclidean distances between pairwise embedding comparisons. By learning to extract meaningful features, the goal is to train the network to map sets of sample arrays belonging to the same subject to points close to each other in the embedding space, whereas embeddings corresponding to different subjects should be as distant as possible.

B. Training Approach

A separate unimodal network is trained for each modality, leading to 10 different models (one for each touch task, one for each background sensor). In fact, during the design phase of the system, it was proved to be beneficial for the system performance to train a single model for each different background sensor data taking sample arrays evenly from the different tasks (i.e., keystroke, scroll up and scroll down, drawing the number "8", tapping on the screen). In such way, the model is able to learn more general and robust features, improving the results in terms of Equal Error Rate (EER) compared to the case of having different models for every *task-modality* combination.

C. Triplet Loss Function

First defined in [27], the triplet loss function enables learning from positive and negative comparisons at the same time. A triplet is composed by three different sample arrays from two different classes: Anchor (A) and Positive (P) are different sample arrays from the same subject, and Negative (N) is a sample array from a different subject. The triplet loss function is defined as follows (1):

$$\mathcal{L}_{TL} = max\{0, d^2(\mathbf{v}_A, \mathbf{v}_P) - d^2(\mathbf{v}_A, \mathbf{v}_N) + \alpha\} \quad (1)$$

where α is a margin between positive and negative pairs and d is the Euclidean distance between pairs consisting in the *anchor* $\mathbf{v}_A$, the *positive* $\mathbf{v}_P$ and the *negative* $\mathbf{v}_N$ embeddings of a given sample array. In a unique operation, triplet loss minimizes the distance between embedding vectors from the same class ($d^2(\mathbf{v}_A, \mathbf{v}_P)$), and maximizes it for embeddings from different classes ($d^2(\mathbf{v}_A, \mathbf{v}_N)$).

D. Triplet Mining Strategy

For every modality, the training is divided into two phases of 50 epochs each.

During the first phase, a model is trained with random triplets. In other words, given a pair of sample arrays belonging to different acquisition sessions of the same subject, namely the *anchor-positive* pair, to complete the triplet, the negative sample array is randomly selected from other subjects' data in the training set. Once the first phase is finished, the checkpoint model of each modality is used in order to compute the embeddings for all the data in the training set.

The second phase consists in training new models from scratch exploiting only semi-hard triplets calculated by the first phase checkpoint model. Given any *anchor-positive* pair, a semi-hard triplet is achieved by mining a negative sample array such that the *anchor-negative* distance in the embedding

space is less than the considered *anchor-positive* distance plus a margin $\alpha = 1.5$. The purpose of triplet mining is training the model with sample arrays that are harder to classify in order to improve the performance of the network with difficult comparisons.

E. Fusion of Modalities

The fusion between different modalities is carried out at score level, which is just one of the possible data fusion strategies [18]. Score level fusion is based on combining the scores, in terms of distances between embeddings, obtained from the different modalities available within the same time window. Among the several possible score level fusion approaches, the approach adopted in the present work is the summation of the scores, as the scores produced by the different networks are in a close range.

During each one of the touch tasks, six modalities are considered: the specific task data (keystroke, scroll up, scroll down, drawing an 8 with the finger, tapping), and the five background sensors (accelerometer, gravity sensor, gyroscope, linear accelerometer, magnetometer), leading to 63 different possible combination subsets.

V. Experimental Protocol

The data subjects are divided into training, validation and test sets without any overlap. The subjects with less than five complete acquisition sessions and with acquisition errors (i.e., arrays of zeros) were discarded. The final number of subjects available depends on the modality (376 per modality, on average). They were divided into 70% for training, 15% for validation and 15% for testing. Regarding the training hyper-parameters, the batch size is 512, the learning rate is set to 0.05, the Adam optimizer is used with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The models were built in `Keras-Tensorflow`.

In order to evaluate the performance of the networks for the purpose of validation and testing, given each subject, their sample arrays are compared to all the other subject sample arrays. Therefore, considering N subjects, there will be $N-1$ impostors (in the final test dataset, $N = 65$ for keystroke, $N = 59$ for scroll up, $N = 58$ for scroll down, $N = 64$ for drawing an 8, and $N = 65$ for tapping). For each subject, sample arrays from the first three acquisition sessions (out of five) are considered as enrollment data, whereas the sample arrays from the remaining 2 sessions are used for verification. At verification time, the score value consists in the mean of the three obtained distance values. Consequently, for every subject 2 genuine distance values will be calculated (one for each test sample array) and $2 \times (N - 1)$ impostor scores.

VI. Experimental Results

During the second phase of the training based on semi-hard triplet mining, three (gyroscope, gravity sensor, linear accelerometer) out of the ten systems developed slightly improved in terms of EER achieved on the validation set ($\sim$1%), with respect to their random triplet counterpart.

The modalities considered can be divided into two categories: *(i)* task data, and *(ii)* background sensor data. The first category refers to a specific task, such as the keystroke data, or the touch data acquired while scrolling. The second category includes data from background sensors acquired during each task.

The results obtained on the HuMIdb test data are presented in Table I. The metric chosen to evaluate the performance of the system is the EER. The EER values for the specific tasks are in general lower than those achieved on all background sensors except for the magnetometer, which consistently proves to be the best performing background sensor. In fact, only with the keystroke and number drawing task data it is possible to obtain lower EER than the magnetometer. In the case of background sensor modalities, the EER is typically included in the 20-30% range. Looking at the touch information, the best performance is achieved for keystroke and then for drawing "8" with 12.2% and 14.3% EER respectively. The other three tasks achieve EER values in the range of 23-25%. This indicates that typing and drawing provide more discriminative information compared to swiping or tapping.

In Table II, the best three subsets originated from the fusion of modalities are included. In any case, the improvement in the performances of the system due to the fusion of modalities is significant. In all tasks, the best five performing subsets of modalities produce an EER value which is around half of the value obtained for the best performing individual modality or task. The best performance is achieved for the task of drawing the number "8" with a fusion of the touch information (finger) and the accelerometer, gyroscope, and magnetometer (F, A, Gy, M), with 5.5% EER. In this case, the EER produced by the fusion of modalities is nearly one third of the one achieved with the touch data only. The second best performance is given by the task of keystroke, combined with the accelerometer, the gyroscope, and the magnetometer (K, A, Gy, M) scores, (i.e., 7.58% of EER).

The touch modalities are present in almost everyone of the best subsets, together with the magnetometer and accelerometer data, whereas the contribution of the other modalities (gravity sensor, gyroscope, and linear accelerometer) is less consistent.

A. Performance Adaptation to Fewer Enrollment Data

A set of experiments is carried out in order to evaluate the ability of the model with fewer data available at verification time. The models are tested with a varying number (1-3) of enrollment sample arrays, and a fixed number (2) of verification sample arrays. The enrollment sample arrays belong to the first 3 acquisition sessions (1, 2, 3), whereas the 2 verification sample arrays always belong to sessions 4 and 5. Considering sample arrays from the first three acquisition sessions as enrollment data and the sample arrays from the last two acquisition sessions as verification data, the values shown consist in the mean of the possible subsets of enrollment sample arrays. Overall, an improving trend as the number of

TABLE I

RESULTS IN TERMS OF EER (%) OF THE DIFFERENT INDIVIDUAL MODALITIES DURING EACH TASK.

	Individual Modalities					
Task	Specific Task	Accelerometer	Gravity Sensor	Gyroscope	Lin. Accelerometer	Magnetometer
Keystroke	**12.19**	26.16	30.81	28.46	27.02	21.36
Scroll up	24.99	29.47	28.27	30.16	25.98	**22.41**
Scroll down	24.36	27.22	25.59	27.33	**21.93**	23.68
Drawing "8"	**14.28**	24.60	26.18	23.80	22.22	19.04
Tap	23.23	23.73	27.11	24.60	23.85	**22.27**

TABLE II

RESULTS IN TERMS OF EER (%) OF THE 3 BEST SUBSETS ORIGINATED FROM THE FUSION OF THE DIFFERENT INDIVIDUAL MODALITIES FOR EACH TASK.

	Fusion of Modalities					
Task	Subset #1	EER (%)	Subset #2	EER (%)	Subset #3	EER (%)
Keystroke	K, A, Gy, M	**7.58**	K, A, M	7.68	K, A, Gr, M	7.81
Scroll up	SU, A, Gy, L, M	**12.93**	SU, A, Gr, Gy, L, M	13.08	SU, A, Gr, M	13.69
Scroll down	A, Gr, L, M	**10.93**	SD, A, L, M	11.40	SD, Gy, L, M	11.40
Drawing "8"	TD, A, Gy, M	**5.47**	TD, A, Gr, Gy, M	5.56	TD, A, Gy, L, M	5.61
Tap	T, A, Gr, Gy, L, M	**10.91**	T, A, Gy, L, M	10.94	T, A, Gr, Gy, M	11.35

Acronyms of Tasks: TD = Touch Draw, K = Keystroke, SD = Scroll Down, SU = Scroll Up, T = Tap. Acronyms of Background Sensors: A = Accelerometer, Gr = Gravity Sensor, Gy = Gyroscope, L = Linear Accelerometer, M = Magnetometer.
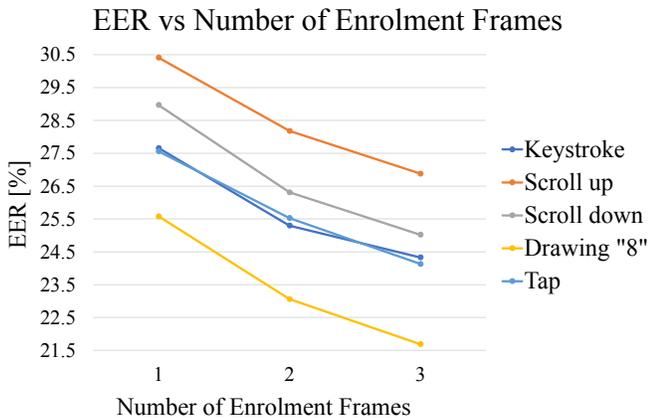

Fig. 1. The improvement trend in terms of EER is consistent across tasks as the amount of enrollment sample arrays increases.

enrollment data increases is noticeable for all tasks. The results are graphically displayed in Fig. 1.

VII. CONCLUSIONS

This paper has provided an analysis of individual behavioral biometric traits suitable for the application of continuous user authentication through the interaction with mobile devices. The modalities considered are based on touchscreen and background sensor data. The dataset considered for our work is the HuMIdb, a novel public multimodal mobile database of sensor data acquired in an unsupervised scenario. We implemented, for each individual modality, a separate LSTM RNN and semi-hard triplet loss, with the fusion of different modalities at score level. The results achieved show that the best performing sources are touch information in keystroke and drawing a number on the screen tasks. Then, the background sensor that provided the best results was the magnetometer in almost all tasks even improving the performance of the main touch information in the cases of scrolling and tapping. In any case, the discriminative ability of the touch modalities is significantly enhanced by the fusion with the available background sensor modalities at score level (typically reaching a 5%-13% EER range).

Other loss functions or different network architectures could be tested in the future. Further analysis focused on an optimal solution for improving the data fusion strategy will be carried out, e.g. more sophisticated frameworks for binary classifier score level fusion [28]. Additionally, starting from the observation that in the real-world scenario of a theft, the impostor and the genuine user data originate from the same device, it would be interesting to assess the device fingerprinting due to sensor differences and calibration imperfections of mobile sensors, and decorrelate user from device recognition in the learned data representation for background sensors [29].

On a separate note, a clear trend of improvement in performance was evident when adding enrollment data, at the cost of a higher memory and computational overload. This aspect could also be further investigated to reach higher security standards for a real-world deployment scenario.

REFERENCES

[1] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens," in *Proc. 4th USENIX Conf. on Offensive Technologies*, 2010.

[2] M. Harbach, E. von Zezschwitz, A. Fichtner, A. De Luca, and M. Smith, "It's a hard lock life: a field study of smartphone (un)locking behavior and risk perception," in *Proc. 10th Symp. On Usable Privacy and Security*, 2014.

[3] E. von Zezschwitz, M. Eiband, D. Buschek, S. Oberhuber, A. De Luca, F. Alt, and H. Hussmann, "On quantifying the effective password space of grid-based unlock gestures," in *Proc. 15th Intl. Conf. on Mobile and Ubiquitous Multimedia*, 2016.

[4] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*. Springer, 2019.

[5] C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, *Handbook Of Digital Face Manipulation And Detection: From DeepFakes to Morphing Attacks*. Springer Nature, 2022.

[6] P. Perera and V. M. Patel, "Efficient and low latency detection of intruders in mobile active authentication," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 6, pp. 1392–1405, 2017.

[7] 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.

[8] A. Acien, A. Morales, J. Fierrez, R. Vera-Rodriguez, and J. Hernandez-Ortega, "Active detection of age groups based on touch interaction," *IET Biometrics*, vol. 8, no. 1, pp. 101–108, 2019.

[9] P. Delgado-Santos, G. Stragapede, R. Tolosana, R. Guest, F. Deravi, and R. Vera-Rodriguez, "A survey of privacy vulnerabilities of mobile device sensors," *ACM Computing Surveys*, 2022.

[10] H. Lu, J. Huang, T. Saha, and L. Nachman, "Unobtrusive gait verification for mobile phones," in *Proc. ACM Intl. Symp. on Wearable Computers*, 2014.

[11] A. Acien, A. Morales, J. V. Monaco, R. Vera-Rodriguez, and J. Fierrez, "Typenet: Deep learning keystroke biometrics," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, pp. 1–1, 2021.

[12] M. Santopietro, R. Vera-Rodriguez, R. Guest, A. Morales, and A. Acien, "Assessing the quality of swipe interactions for mobile biometric systems," in *Proc. IEEE/IAPR Intl. Joint Conf. on Biometrics (IJCB)*, 2020.

[13] R. Tolosana, J. C. Ruiz-Garcia, R. Vera-Rodriguez, J. Herreros-Rodriguez, S. Romero-Tapiador, A. Morales, and J. Fierrez, "Child-computer interaction with mobile devices: Recent works, new dataset, and age detection," *IEEE Trans. on Emerging Topics in Computing*, pp. 1–1, 2022.

[14] R. Tolosana *et al.*, "SVC-onGoing: Signature verification competition," *Pattern Recognition*, 2022.

[15] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Reducing the template ageing effect in on-line signature biometrics," *IET Biometrics*, vol. 8, no. 6, pp. 422–430, 2019.

[16] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Biometric signature verification using recurrent neural networks," in *Proc. 14th IAPR Int. Conference on Document Analysis and Recognition, ICDAR*, 2017.

[17] Z. Akhtar, A. Hadid, M. S. Nixon, M. Tistarelli, J.-L. Dugelay, and S. Marcel, "Biometrics: In search of identity and security (QA)," *IEEE MultiMedia*, vol. 25, no. 3, pp. 22–35, 2018.

[18] M. Singh, R. Singh, and A. Ross, "A comprehensive overview of biometric fusion," *Information Fusion*, vol. 52, pp. 187–205, 2019.

[19] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognition Letters*, vol. 79, pp. 80–105, 2016.

[20] M. Abuhamad, A. Abusnaina, D. Nyang, and D. Mohaisen, "Sensor-based continuous authentication of smartphones' users using behavioral biometrics: a contemporary survey," *IEEE IoT Journal*, vol. 8, no. 1, pp. 65–84, 2021.

[21] A. Acien, A. Morales, R. Vera-Rodriguez, J. Fierrez, and R. Tolosana, "Multilock: Mobile active authentication based on multiple biometric and behavioral patterns," in *Proc. ACM Intl. Conf. on Multimedia, Workshop on Multimodal Understanding and Learning for Embodied Applications*, 2019.

[22] D. Deb, A. Ross, A. K. Jain, K. Prakah-Asante, and K. V. Prasad, "Actions speak louder than (pass)words: Passive authentication of smartphone* users via deep temporal features," in *Proc. 2019 Intl. Conf. on Biometrics*, 2019.

[23] A. Acien, A. Morales, R. Vera-Rodriguez, and J. Fierrez, "Smartphone sensors for modeling human-computer interaction: General outlook and research datasets for user authentication," in *Proc. IEEE Intl. Workshop on Consumer Devices and Systems*, 2020.

[24] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: a unified embedding for face recognition and clustering," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2015.

[25] A. Acien, A. Morales, J. Fierrez, R. Vera-Rodriguez, and O. Delgado-Mohatar, "BeCAPTCHA: Behavioral bot detection using touchscreen and mobile sensors benchmarked on HuMIdb," *Engineering Applications of Artificial Intelligence*, vol. 98, p. 104058, 2021.

[26] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *arXiv:1506.00019*, 2015.

[27] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, no. 9, pp. 207–244, 2009.

[28] N. Brümmer and E. de Villiers, "The bosaris toolkit: Theory, algorithms and code for surviving the new dcf," 2013.

[29] 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.