Multi-Modality Mobile Datasets for Behavioral Biometrics Research

Data/Toolset paper

Aratrika Ray-Dowling
Department of Electrical and Computer Engineering
Clarkson University
Potsdam, NY, USA
ray@clarkson.edu

Ahmed Anu Wahab
Department of Electrical and Computer Engineering
Clarkson University
Potsdam, NY, USA
wahabaa@clarkson.edu

Daqing Hou
Department of Electrical and Computer Engineering
Clarkson University
Potsdam, NY, USA
dhou@clarkson.edu

Stephanie Schuckers
Department of Electrical and Computer Engineering
Clarkson University
Potsdam, NY, USA
sschucke@clarkson.edu

ABSTRACT

The ubiquity of mobile devices nowadays necessitates securing the apps and user information stored therein. However, existing one-time entry-point authentication mechanisms and enhanced security mechanisms such as Multi-Factor Authentication (MFA) are prone to a wide vector of attacks. Furthermore, MFA also introduces friction to the user experience. Therefore, what is needed is continuous authentication that once passing the entry-point authentication, will protect the mobile devices on a continuous basis by confirming the legitimate owner of the device and locking out detected impostor activities. Hence, more research is needed on the dynamic methods of mobile security such as behavioral biometrics-based continuous authentication, which is cost-effective and passive as the data utilized to authenticate users are logged from the phone's sensors. However, currently, there are not many mobile authentication datasets to perform benchmarking research. In this work, we share two novel mobile datasets (Clarkson University (CU) Mobile datasets I and II) consisting of multi-modality behavioral biometrics data from 49 and 39 users respectively (88 users in total). Each of our datasets consists of modalities such as swipes, keystrokes, acceleration, gyroscope, and pattern-tracing strokes. These modalities are collected when users are filling out a registration form in sitting both as genuine and impostor users. To exhibit the usefulness of the datasets, we have performed initial experiments on selected individual modalities from the datasets as well as the fusion of simultaneously available modalities.

CCS CONCEPTS

• **Security and privacy → Biometrics**.

CODASPY '23, April 24–26, 2023, Charlotte, NC, USA

ACM ISBN 979-8-4007-0067-5/23/04…$15.00
https://doi.org/10.1145/3577923.3583637

KEYWORDS

Mobile Datasets, Continuous Authentication, Multi-Modalities, Fusion, Behavioral Biometrics

ACM Reference Format:
Aratrika Ray-Dowling, Ahmed Anu Wahab, Daqing Hou, and Stephanie Schuckers. 2023. Multi-Modality Mobile Datasets for Behavioral Biometrics Research: Data/Toolset paper. In *Proceedings of the Thirteenth ACM Conference on Data and Application Security and Privacy (CODASPY '23), April 24–26, 2023, Charlotte, NC, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3577923.3583637

1 INTRODUCTION

Currently, there are billions of cell phone users whose sensitive information such as emails, online transactions, messages, multimedia files, and other data is stored in their mobile devices. Hence, to protect such sensitive information, traditional one-time, entry-point security mechanisms (e.g., PINs, passwords, patterns, and biometrics) are not secure enough as they can be compromised. On the other hand, MFA, being a more advanced security mechanism, introduces friction in the user experience as users need to perform extra tasks, e.g., typing one-time passwords. To overcome such disadvantages we need a continuous, non-intrusive, and dynamic security mechanism. Behavioral Biometrics based continuous authentication, in addition to being non-intrusive and dynamic, is also cost-effective since the data required for user authentication can be logged from the phone's sensors. Utilizing such a mechanism we can authenticate the legitimate user of the device and lock out impostor activities.

There currently are not enough publicly available mobile datasets to benchmark algorithms [6, 9] to advance the state of the art. In this work, we present two novel mobile datasets with multiple behavioral biometric modalities, namely, swipe, keystroke, acceleration, gyroscope, rotation, and stroke. Our datasets are collected when users are filling out a registration form in a seated posture, which is a common activity. The seated posture of the users makes the scenario more challenging as the magnitudes of the non-sporadic motion sensors (acceleration, gyroscope, and rotation) are not high when users are stationary. In such instances, when other stronger

yet sporadic modalities (swipe and keystroke) are absent, the legitimate user of the device can still be reliably authenticated through the motion sensors. Such reliability of our data is further supported by the results of our initial experiments.

When a user interacts with a mobile device, they tend to achieve stability and precision. Users also attend a postural preference. The user's stability, precision, and postural preference, along with physiological traits (muscular flexing, arm size, grip strength, and age) together contribute to the unique characteristics of users [8]. Such uniqueness can be measured through the motion sensors (acceleration, gyroscope, and rotation) logged continuously when users sit and fill out the registration form. Additionally, we also log swipes and keystrokes during the form-filling procedure which can be utilized either individually or by fusing with the non-sporadic motion sensors. These datasets will be shared upon request.

The Android application through which we log multiple modalities consists of a registration form that includes three routine usage scenarios, namely, filling the form using a user's own credentials, using the fingerprint scanner, and tracing geometrical patterns along guided outlines. The form is a prototype of registration forms required in online account opening scenarios. Fingerprint scanning is a novel behavior during which a user is authenticated only through the motion events logged when the scanning hardware is utilized. Such security measure is useful when the user's phone is stolen by an impostor (e.g., a trusted friend, family member, or colleague) who is capable of bypassing the entry-point security and spoofing the fingerprint reader to get access to applications that are locked using the genuine user's fingerprint (banking or shopping apps). Lastly, pattern tracing is a prototype of unlocking phones using patterned passwords. Users are authenticated through the strokes logged during shape tracing to provide an additional security measure. In this case, beyond knowing the right pattern, the user needs to be authenticated through strokes.

There are few existing mobile behavioral biometric datasets, as shown in Table 1, where our datasets involve three novel user behavior scenarios. No other datasets so far include the above behavior scenarios for seated users.

We have tested the usefulness of our datasets by performing authentication experiments on selected modalities. We perform experiments on swipe, stroke, acceleration, and gyroscope as individual modalities and fusion of motion events from the CU Mobile dataset I. Furthermore, we perform authentication experiments on keystrokes from the CU Mobile dataset II.

The rest of the paper is presented as follows. Section 2 discusses the related work. Section 3 describes the two datasets. Section 4 presents the initial experiments performed on the selected modalities from the datasets. Lastly, Section 5 concludes our paper.

2 RELATED WORK

This section describes related work on the currently available public Mobile Behavioral Biometric datasets as shown in Table 1.

The public dataset BB-MAS, by Belman et al. (2019) [1], includes 117 recruited volunteers, providing data on three different interfaces (phone, tablet, and desktop) while performing multiple activities. The dataset includes acceleration, gyroscope, swipes, keystrokes, and mouse event modalities.

Table 1: Review of publicly available mobile datasets

Dataset	#users	Modalities	User Behavior	Best perf.
BB-MAS [1]	117	accel, gyro, key, swipe, mouse	typing, gait in multiple postures	None
HMOG [8]	100	accel, gyro, magneto, key, swipe, tap, pinch	read/write/map (sit and walk)	7.16%, EER
BrainRun [5]	2,218	accel, gyro, magneto, swipe, tap	game playing	None
UMDAA-02 [3]	48	accel, gyro, magneto, face, touch, light, temp, prox, GPS, BT, WiFi, pressure	routine usage	96.6%, Accur.
Touchalytics [2]	41	swipe	swiping in sitting	0%, EER
LTU Touch [7]	190	swipe	swiping in sitting	10%, EER
BehavePassDB [9]	81	accel, gyro, magneto, lin-accel, gravity, prox, light, temp, pressure, humid, GPS, WiFi, BT, battery	routine usage	87.2%, AUC
CU Mobile Datasets I & II	**88**	**accel, gyro, rotate, swipe, double tap, key, touch, stroke**	**form filling in sitting**	**2.4%, EER**

The HMOG public dataset by Sitová et al. (2015) [8] has 100 recruited participants where the data is collected on mobile devices. Users provide data in 24 sessions by performing reading, writing, and map navigation while sitting and walking. Out of the 8 typing/writing sessions, there are four sessions that require users to sit and type. The dataset includes acceleration, gyroscope, magnetometer, swipe, tap, pinch, and keystroke modalities.

The large public dataset BrainRun by Papamichail et al. (2019) [5] includes 2,218 volunteers who perform gaming activities while accelerometer, gyroscope, magnetometer, swipe, and tap are logged.

The UMDAA-02 dataset by Mahbub et al. (2016) [3] involves routine phone usage of 48 volunteers through which they log from the front-facing camera, touchscreen, gyroscope, accelerometer, magnetometer, light sensor, GPS, Bluetooth (BT), WiFi, proximity, temperature, and pressure sensors.

The Touchalytics dataset by Frank et al. (2013) [2] includes single modality swipe data from 41 volunteers. Users participate in three sessions to provide horizontal and vertical swipes.

Another similar dataset that involves single swipe modality, is the LTU Touch dataset by Serwadda et al. (2013) [7], where 190 recruited volunteers swipe across multiple choice questions.

In a recent study by Stragapede et al. (2022) [9], a routine phone usage-based dataset is developed utilizing readings from 15 different phone sensors. The study performs both single and multi-modality experiments where the fusion of modalities improves the system performance. Their work also involves impostor attack scenarios.

Our pair of novel mobile datasets involve three real-life usage scenarios, namely, registration form filling, fingerprint scanning, and tracing geometrical shapes during which we capture multiple modalities from 88 seated users. Authentication during the sitting posture will mostly depend on motion events when sporadic non-motion modalities are not available. While sitting and typing are common across existing datasets [1, 8, 9], the behaviors of filling out forms, using the fingerprint scanner, and tracing shapes are

Table 2: User demographics in CU Mobile datasets.

Demographics	Type/ Group	CU dataset I	CU dataset II
Age	18-20	17	17
(increasing order)	21-25	9	11
	26-30	12	9
	31-35	5	0
	>35	6	2
Gender	Female	23	20
(alphabetical order)	Male	26	19
Ethnicity	Asian	16	8
(alphabetical order)	Black or African American	6	7
	Hispanic or Latino	2	3
	White	23	22
	UNKNOWN	3	1

unique to ours. Additionally, the datasets will benefit the research community to perform benchmarking research in continuous authentication on mobile devices when users are performing three different real-life usage scenarios. Our datasets are a contribution to advance the current state of the art when at present there are not many mobile datasets. Additionally, our number of users is comparable to the few publicly available datasets [1, 3, 8, 9].

3 CU MOBILE DATASETS I & II

3.1 Data collection

The collections of CU Mobile datasets I and II are approved by Clarkson University's IRB (Institutional Review Board). The only criterion of volunteer recruitment is the age requirement of being 18 years old or above. Each user visits twice with a gap of at least two days between the visits. The durations of the first and second visits are limited to 1 hour and 1.5 hours respectively. In the first visit, each user uses the Android logger 10 times where they fill out the form with their own information. In the second session, each user uses the logger 15 times where in the first 5 times they fill out the form with their own information, and the last 10 times they perform an attack scenario by filling out the form with the data of 5 previous volunteers where each user's data is typed twice. During the attack scenario, users are not told whether the flashcards contain other participants' information. They are asked to enter the information as perfectly as possible. Furthermore, the users are prompted before they start providing data to not use their personal PINs or passwords for the form's password field and to make up one on the spot. For all other fields, they utilize their own information. Data is collected in a lab environment. The only differences of ours from outside, natural environments are the scheduled time of the data collection and the presence of the experimenter to ensure the smoothness of the collection. We ask users to sit comfortably without restricting the admission of others into the lab or controlling the level of noise.

The CU Mobile dataset I, collected in 2019, includes multi-modality data from 49 volunteers. On the other hand, the CU Mobile dataset II, collected in 2020, following COVID-19 safety protocols, includes multi-modality data from 39 volunteers. The demographic information of the participants is shown in Table 2.

From both datasets, we could collect multi-modality data from 88 volunteers in total. The keystroke data in CU Mobile dataset I does not have key release timestamps. We also observe that in the CU Mobile dataset I, volunteers made spelling errors for the same input to a text field, between iterations. Therefore, to overcome the above two issues we collect CU Mobile dataset II whose keystroke data has both key press and release timestamps. Additionally, it has a checker at the application level which compares entries in the fields across iterations. If an error occurs beyond a minimum threshold then the users are notified about the spelling errors between iterations. Thereafter users need to retype the iteration to enter the error-free information. Therefore, in the future, researchers may individually utilize the datasets or may combine the 88 users' data for all modalities except the keystroke. The sampling rate of each of the motion events is 2 Hz in CU Mobile dataset I. While collecting the first dataset we observe that the motion events (acceleration, gyroscope, and rotation) are allowed by the API to be logged at a higher sampling rate than the rate specified in the Android application. Hence, in the second dataset we vary the sampling rates per motion event within the values of 2, 5, and 10 Hz.

3.2 Android logger

The Android application involves three routine usages, namely, filling out the form, using the fingerprint scanner, and tracing geometrical patterns along guided outlines. The logger captures keystrokes, swipes, motion events (acceleration, gyroscope, and rotation), motion events during fingerprint scanning, and strokes from pattern tracing. We also log other on-touch events, namely, taps, double taps, and longpresses in addition to swipes. However, we observe few samples of longpress per user. Also, there are users who do not exhibit longpress at all. The occurrence of double taps is mostly triggered when some users by habit double click on edit-text fields to initiate typing during form filling. The single touch/ tap events have samples that are subsets to other touch events (keystroke, swipe, double tap, and longpress). However, we have independent touch events which occur when a user taps on the screen beyond inputting other touch events.

The registration form consists of the following fields: *Full name, Email, Phone, Address, City, State, Zip, Password, Age, and Declaration statement.* We utilize a virtual keyboard to log keystroke data from users. Figure 1 shows the Android logger.

Android provides an API to read data from the phone's embedded motion sensors (accelerometer, gyroscope, and rotation). The sensor listeners read data from the sensor at pre-defined time intervals. The collected motion events during both form-filling and fingerprint-scanning activities are separated in our database. The fingerprint sensor is called when the user presses *INPUT FINGERPRINT* button. It is designed to capture the fingerprint for 5 times. Once fingerprint reading is done, the user clicks the *REGISTER* button after which they are asked to trace regular shapes (triangle, square, hexagon, octagon, 5-point star, and 6-point star) along a guided outline.

The CU Mobile dataset I is stored in SQLite3 database files and the CU Mobile dataset II is stored in MySQL server which is extracted to CSV files and can be shared upon signing a release form.

3.3 Mobile devices used

During the collection of the CU Mobile dataset I we utilize the following Android devices: Samsung Galaxy S8 (size: 5.8 inches), Samsung Galaxy Note 9 (size: 6.4 inches), and Motorola X4 (size: 5.2 inches). Each user is provided with a Samsung Galaxy S8 and

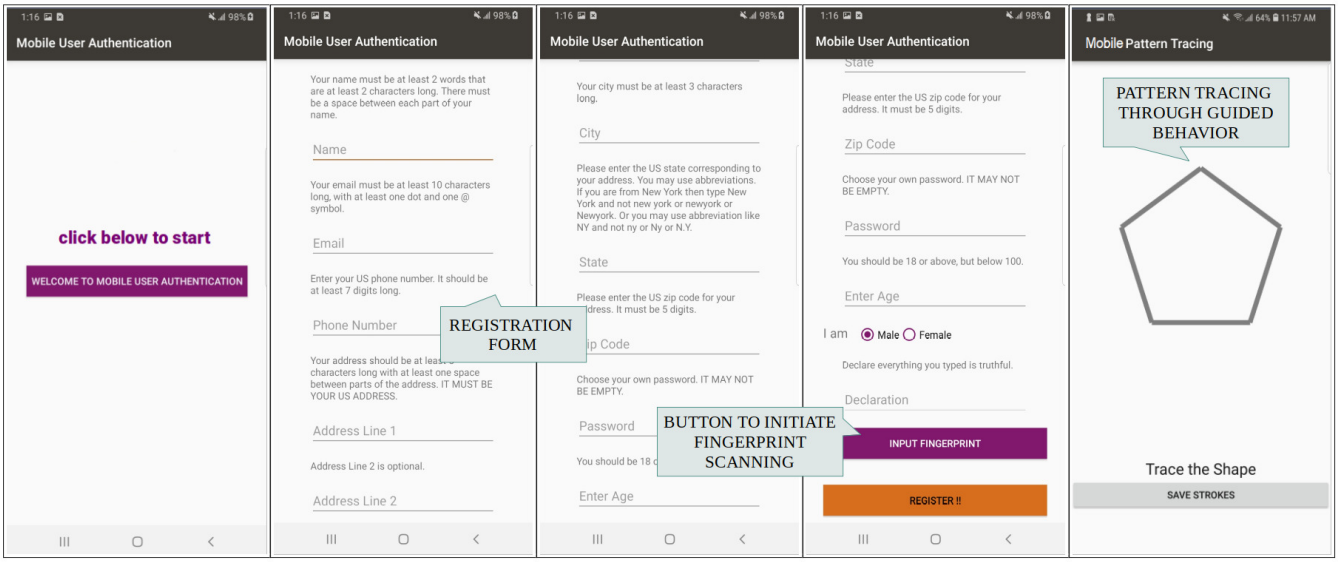

Figure 1: User interface of our Android logging app where a user simulates the account recovery process by filling out a form with personal information to reclaim an account, performing fingerprint scanning, and tracing geometrical patterns.

any one of Motorola X4 and Samsung Galaxy Note 9. Therefore, this dataset is collected using two mobile devices per user in each visit. All of these devices have inbuilt sensors like an accelerometer, gyroscope, magnetometer, and fingerprint. Collecting data on two different devices will enable future researchers to explore the potential impact of devices on performance. During the collection of CU Mobile dataset II, we only utilize Samsung Galaxy S8.

3.4 Multiple modalities

Our Android application lets users interact with the mobile device using three usage scenarios. During each scenario, we collect different modalities which are listed as follows.

- Registration form filling: keystroke, touch/ tap, double tap, longpress, swipe, acceleration, gyroscope, and rotation.
- Fingerprint scanning: acceleration, gyroscope, and rotation.
- Pattern tracing: stroke (smaller swipe segments).

Table 3 shows the data columns logged per modality and the raw data statistics. For CU Mobile dataset I, the statistics per modality is calculated based on the cumulative data collected from the two phones given to each user. In CU Mobile dataset II there is no such scenario since data is collected from only one mobile device.

3.5 Data privacy & data provision

To protect the personally identifiable information (PII) in our keystroke data, the "key" column, was encrypted using the substitution cipher method. With an encryption key, the substitution cipher consistently replaces each character of the plaintext with another letter, number, or symbol.

Although there are many other encryption methods available, we choose the substitution cipher due to the keystroke dynamics authentication mechanism where instances of each graph in a test sample are compared with the graph in the profile samples. Hence the encryption of each character must be consistent (e.g., if the cipher of "a" is "w", then all "a" must be substituted with the ciphertext "w"). In this way, the encrypted data remains usable for keystroke dynamics research and can be released to a third party in CSV format upon signing a release form.

A known drawback of the substitution cipher is that it can be broken with frequency analysis. Frequency analysis relies on having a sufficiently large block of cipher text to create distributions of individual letters and n-grams that match the statistical profile of a language (e.g., English). To overcome this method of cryptanalysis, we encrypt each field of each user with a different key, thereby breaking the connection between fields. That is, no two fields share the same encryption key. Therefore no reliable distributions of n-grams can be formed with the short fields. This approach should mitigate against all possible manual cryptanalysis and should suffice in protecting the PII in our data. Furthermore, a license agreement will be put in place to further ensure that only authorized researchers may use our datasets for research purposes. To further enhance the usefulness of the dataset, we add a column indicating whether the encrypted key is a digit, alphabet, or symbol.

4 INITIAL EXPERIMENTS

We perform experiments on selected modalities from both datasets. We demonstrate both single and multi-modality experiments. The performance metric utilized across all the experiments is EER (Equal Error Rate) which occurs when the False Acceptance Rate (FAR) is equal to the False Rejection Rate (FRR) on the ROC curve.

4.1 Experiments on keystrokes

We pre-process the raw keystroke data by removing backspaces and the keystrokes deleted by backspaces, which have been used for correcting misspellings. Thereafter we extract the down-down (DD)

Table 3: Data columns per modalities collected from phone's sensors and statistics of the datasets

Modality	Columns	CU I data statistics (avg/med/min/max/stdv)	CU II data statistics (avg/med/min/max/stdv)
Swipe	timestamps (ms), x coordinate (px), y coordinate (px), pressures ($[0, 1]$), orientations, sizes ($[0, 1]$)	254 /280 /69 /423 /87	253/ 256/ 89/ 496/ 103
Double Tap	timestamps (ms), x coordinates (px), y coordinates (px)	1650 /1610 /95 /3184 /642	1394/ 1530/ 393/ 2401/ 489
Keystroke	field, key, x coordinate (px), y coordinate (px), press timestamp (ms), release timestamp (ms)	5267 /5307 /1710 /13317 /1943	4259/ 4699/ 1473/ 6376/ 1334
Acceleration	timestamp (ns), x-acceleration (m/s^2), y-acceleration (m/s^2), z-acceleration (m/s^2)	37373 /26640 /3687 /139919 /33391	14677/ 10284/ 1972/ 94411/ 16297
Gyroscope	timestamp (ns), x-gyroscope (rad/s), y-gyroscope (rad/s), z-gyroscope (rad/s)	6241 /6282 /1879 /13096 /2393	17667 / 6147/ 1397/ 135970/ 25970
Rotation	timestamp (ns), x-rotation (rad), y-rotation (rad), z-rotation (rad)	17134 /16839 /5417 /38020 /6859	19361 / 20167/ 5122/ 38809/ 7594
Fingerprint Accel	timestamp (ns), x-acceleration (m/s^2), y-acceleration (m/s^2), z-acceleration (m/s^2)	4053 /2103 /311 /21586 /4711	463/ 324/ 53/ 1679/ 428
Fingerprint Gyro	timestamp (ns), x-gyroscope (rad/s), y-gyroscope (rad/s), z-gyroscope (rad/s)	556 /460 /155 /2128 /345	437/ 198/ 53/ 2567/ 561
Fingerprint Rotate	timestamp (ns), x-rotation (rad), y-rotation (rad), z-rotation (rad)	1438 /1316 /411 /3834 /701	574/ 599/ 140/ 1399/ 252
Stroke	shape type, shape sub-type, timestamp (ms), x coordinate (px), y coordinate (px)	346 /311 /111 /2148 /309	107/ 100/ 29/ 385/ 59

Table 4: Authentication based on individual fields for the keystroke dataset. The best performances are shown in bold.

Field	Zip	City	Phone	Name	Address	Email	Declare
# Avg shared diagraph	3	6	6	12	16	14	43
EER (%)	32.17	20.86	25.13	**15.29**	**13.1**	19.42	**12.26**

features from the data. We apply the Scaled-Manhattan distance algorithm on individual fields and perform weighted score fusion. We treat each field individually and compare only the profile and test samples of the same field. The weighted score fusion is the weighted sum of individual field scores d_i, where all weights sum up to one (see Equations 1 and 2). We use grid search method [8] to find the optimum weights for each combination, with the minimum weight being 0.05 and an increment of 0.05 at every search iteration.

$$D = w_1 \times d_1 + w_2 \times d_2 + ... + w_N \times d_N \quad (1)$$

$$w_1 + w_2 + ... + w_N = 1 \quad (2)$$

The results of the individual fields are shown in Table 4. "Declare", "Address" and "Fullname" are the three best-performing individual fields with an EER of 12.36%, 13.1%, and 15.29% respectively. Among other possible reasons such as familiarity, the results suggest that keystroke length can significantly influence performance.

With fusion technique, there are 21 possible combinations for Duet (two fields), 35 for Trio (three fields) and Quartet (four fields), 21 for Quintet (five fields), 7 for Sextet (six fields), and 1 for Septet (seven fields). Table 5 shows the results for the best weighted score fusion combinations. As the field combination increases from Duet (two fields) to Septet (seven fields), performance improves accordingly, showing that more data (longer text) results in better performance. The overall best performance of 2.64% EER is achieved at the combination of seven fields (Septet) [10].

4.2 Experiments on motion events

We perform both single and multi-modalities experiments taking acceleration and gyroscope (of CU Mobile dataset I) logged during form filling. We perform weighted score level fusion and Nandakumar et al.'s (2006) Likelihood Ratio (LR) based score fusion [4] to fuse multiple modalities (acceleration and gyroscope). These

Table 5: Weighted score fusion of multiple fields for the keystroke data, w =weight. Best performance is shown in bold.

	Best Combinations	EER (%)
DUET	Declare (w=0.75)+Address (w=0.25)	5.28
TRIO	Email (w=0.45)+Declare (w=0.25)+ Fullname (w=0.3)	5.14
QUARTET	Fullname (w=0.4)+Address (w=0.35)+ City (w=0.15)+Phone (w=0.1)	3.65
QUINTET	Email (w=0.1)+Declare (w=0.2)+ Fullname (w=0.4)+ Address (w=0.15)+City (w=0.15)	3.44
SEXTET	Email (w=0.2)+Declare (w=0.3)+ Fullname (w=0.2)+Address (w=0.15)+ City (w=0.05)+Zip(w=0.1)	2.81
SEPTET	Email (w=0.2)+Declare (w=0.25)+ Fullname (w=0.2)+Address (w=0.2)+ City (w=0.05)+Zip (w=0.05)+Phone (w=0.05)	**2.64**

multi-modalities experiments ensure the feasibility of the fusion of simultaneously available modalities.

For single modality (acceleration or gyroscope) we fuse k distance scores to enhance the authentication performance. The sliding window with k distance scores slides by n stride. We split each user's data into training, testing, and validation. We validate the k and n values, where k ranges from 5 to 150 with step size 5 and n ranges from 5, 10 to 140 with step size 10.

For weighted score level fusion, weights are assigned to each motion event, which is multiplied by the generated scores from their respective classifiers, to perform a weighted sum. The resultant sum becomes a new score, k of which is fused for enhanced performance. In addition to k and n, we also validate weight per modality in the range between 0.0 to 1.0 with step size 0.1. The sum of every pair of acceleration and gyroscope weight is always 1. For LR-based fusion, we take 2-dimensional vectors of match scores of acceleration and gyroscope from their respective SVMs. We create genuine and impostor distributions which are estimated as Gaussian Mixture Models (GMM). The LR is defined as the ratio of genuine to impostor distribution. For LR, we also validate the number of Gaussian components ranging from 2 to 18 [6].

We perform both intra-session (training, testing, and validation from visit-1 data per user) and inter-session (training and validation

Table 6: Experimental results of single modality authentication (acceleration and gyroscope). Bold-faced are the best results obtained.

Single modality experiments	Intra-session EER (%) avg/med/min/max/std	Inter-session EER (%) avg/med/min/max/std
Acceleration	20.5/ 20.2/ 3/ 35/ 7	**8.4/ 4.1/ 0/ 31/ 10.8**
Gyroscope	18.3/ 19.9/ 0/ 34.6/ 9.6	**8.5/ 5.7/ 0/ 34.9/ 10.9**

Table 7: Experimental results for fusion of acceleration and gyroscope from Mobile dataset I. Bold-faced are the best results obtained.

Cross-modality experiments	Intra-session EER (%) avg/med/min/max/std	Inter-session EER (%) avg/med/min/max/std
Weighted Score	8.3/ 8/ 0/ 28/ 5.6	7.9/ 0.8/ 0/ 34.5/ 11.5
Likelihood Ratio	**2.4/ 0.9/ 0/ 15.3/ 3.3**	**6.9/ 2/ 0/ 33.4/ 10.1**

Table 8: Experimental results using pattern tracing strokes.

Shape	Triangle	Square	Hexagon	Octagon	5-point start	6-point star
EER (%)	**5**.7	8.5	9.4	14.4	14.2	16.5

from visit-1 and testing from visit-2 per user) experiments. We utilize a Support Vector Machine (SVM) for each experiment where the SVM parameters are set to Radial Basis Function (RBF) kernel, $C = 100$, and $gamma = auto$. Each genuine user is authenticated against all other impostors. The average EER obtained across all genuine users is the overall performance. Tables 6 and 7 show the results of single and cross-modalities experiments respectively.

4.3 Experiments on swipes

We perform experiments on swipes from the CU Mobile dataset I. We extract Frank et al.'s (2013) Touchalytics features [2] on the swipes. Utilizing binary SVM, we perform both intra-session (both training and testing from visit-1 data per user) and inter-session (training from visit-1 and testing from visit-2) experiments. Here, each genuine user is authenticated against all other impostors. The SVM parameters are set to RBF kernel, $C = 100$, and $gamma = auto$. We fuse k distance scores from the SVM to enhance the authentication result, where we set $k = 12$ as the sliding window value. The average EER obtained across all genuine users estimates the overall performance. For the intra-session experiment, the average EER obtained is 1% and for the inter-session experiment, the average EER obtained is 10%.

4.4 Experiments on pattern tracing strokes

We utilize the stroke data during pattern tracing from the CU Mobile dataset I. Here we perform binary classification using SVM where the same classifier parameters are retained as in the swipe experiments (Section 4.3). We perform experiments per shape data, where in each experiment, we fuse $k = 12$ distance scores for better performance. Each genuine user is authenticated against all other impostors. For the overall estimation of performance, we average the EERs across users. The results of experiments performed per shape are shown in Table 8. We observe an increase in the EER with the increased shape complexity.

5 CONCLUSION

We present two novel behavioral biometrics-based mobile datasets (CU datasets I and II) consisting of a total of 88 users. Both are multi-modality datasets with data logged from keystroke, swipe, motion events, and stroke. The data is collected while seated users are performing three common behaviors, namely, filling out a registration form, using the fingerprint scanner, and tracing shapes. No other mobile datasets include these behaviors. We perform initial experiments on selected modalities from the two datasets which demonstrate the usefulness of the data. We also show the feasibility of fusion-based experiments. Given the lack of publicly available mobile datasets, our datasets can be useful to the behavioral biometrics community to enhance research on mobile security.

ACKNOWLEDGMENTS

This material is based upon work supported by CITeR (Center for Identification Technology Research) and the National Science Foundation under Grant No. 1650503. Hou, Schuckers, and Wahab are also partially supported by NSF Award TI-2122746. We are thankful to Robert Newman a.k.a. Bobby for providing the code of the geometric pattern interface portion of the Android application based on which we have developed the present version.

REFERENCES

[1] Amith K Belman, Li Wang, SS Iyengar, Pawel Sniatala, Robert Wright, Robert Dora, Jacob Baldwin, Zhanpeng Jin, and Vir V Phoha. 2019. Insights from BB-MAS–A Large Dataset for Typing, Gait and Swipes of the Same Person on Desktop, Tablet and Phone. *arXiv:1912.02736* (2019).

[2] Mario Frank, Ralf Biedert, Eugene Ma, Ivan Martinovic, and Dawn Song. 2012. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE transactions on information forensics and security* 8 (2012).

[3] Upal Mahbub, Sayantan Sarkar, Vishal M Patel, and Rama Chellappa. 2016. Active user authentication for smartphones: A challenge data set and benchmark results. In *2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS)*. IEEE.

[4] Karthik Nandakumar, Yi Chen, Sarat C Dass, and Anil Jain. 2007. Likelihood ratio-based biometric score fusion. *IEEE transactions on pattern analysis and machine intelligence* 30 (2007).

[5] Michail D Papamichail, Kyriakos C Chatzidimitriou, Thomas Karanikiotis, Napoleon-Christos I Oikonomou, Andreas L Symeonidis, and Sashi K Saripalle. 2019. BrainRun: A behavioral biometrics dataset towards continuous implicit authentication. *Data* 4 (2019).

[6] Aratrika Ray-Dowling, Daqing Hou, Stephanie Schuckers, and Abbie Barbir. 2022. Evaluating multi-modal mobile behavioral biometrics using public datasets. *Computers & Security* 121 (2022).

[7] Abdul Serwadda, Vir V Phoha, and Zibo Wang. 2013. Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms. In *2013 IEEE sixth international conference on biometrics: theory, applications and systems (BTAS)*. IEEE.

[8] Zdeňka Sitová, Jaroslav Šeděnka, Qing Yang, Ge Peng, Gang Zhou, Paolo Gasti, and Kiran S Balagani. 2015. HMOG: New behavioral biometric features for continuous authentication of smartphone users. *IEEE Transactions on Information Forensics and Security* 11 (2015).

[9] Giuseppe Stragapede, Ruben Vera-Rodriguez, Ruben Tolosana, and Aythami Morales. 2022. BehavePassDB: Public Database for Mobile Behavioral Biometrics and Benchmark Evaluation. *Pattern Recognition* (2022).

[10] Ahmed Anu Wahab, Daqing Hou, Stephanie Schuckers, and Abbie Barbir. 2022. Securing account recovery mechanism on desktop computers and mobile phones with keystroke dynamics. *SN Computer Science* 3, 5 (2022), 360.