IET Biometrics

Review Article

IET Journals
The Institution of Engineering and Technology

Data behind mobile behavioural biometrics – a survey

ISSN 2047-4938
Received on 31st May 2019
Revised 27th January 2020
Accepted on 25th February 2020
E-First on 12th October 2020
doi: 10.1049/iet-bmt.2018.5174
www.ietdl.org

Teodors Eglitis[1] ✉, *Richard Guest*[1], *Farzin Deravi*[1]

[1]*School of Engineering and Digital Arts, University of Kent, Kent, Canterbury CT2 7NT, UK*

✉ *E-mail: T.Eglitis@kent.ac.uk*

Abstract: Behavioural biometrics is becoming more and more popular. It is hard to find a sensor that is embedded in a mobile/wearable device, which cannot be exploited to extract behavioural biometric data. In this study, the authors give the reader an overview of mobile device behavioural biometric data and how this data is used in experiments, especially examining papers that introduce new datasets. They will not examine performance accomplished by the algorithms used since a system's performance is enormously affected by the data used, its amount and quality. Altogether, 40 papers are examined, assessing how often they are cited, have databases published, what modality data are collected, and how the data is used. They offer a roadmap that should be taken into account when designing behavioural data collection and using collected data. They further look at the General Data Protection Regulation, and its significance to the scientific research in the field of biometrics. It is possible to conclude that there is a need for publicly available datasets with comprehensive experimental protocols, similarly established in facial recognition.

1 Introduction

Biometric technology is getting more and more popular and accepted by society, mainly due to its success in mobile devices – a wide range of traditional biometric modalities are used in modern smartphones, including fingerprints, iris and face. This is unsurprising – the Bank of America 'Trends in Consumer Mobility Report 2016' found that on an average day 39% of millennials interact with their smartphone more than anything or anyone else and feel anxious when they do not have access to their smartphone [1]. Even without an alert from the mobile device, we decide that we 'must' check in on social media – and thus our phones – immediately [2].

Today, the idea of the use of behavioural biometrics does not seem too unusual. Even South Park – the animated TV series jokes about this topic. In a 2016 episode of the satire show titled 'Fort Collins', one of the characters invents 'emoji analysis', which allows an individual to figure out each student's and teacher's emoji usage and compare it to the person who is *trolling* the other characters [3].

There is an intersection between mobile behavioural biometrics and cognitive psychology because many of the features exploited in behavioural biometrics can be seen as sequences of motor actions [4]. Thus, there should be some lessons learned from the way humans learn to the possible effect on behavioural biometrics. In Psychology, it is known as the 'power law of practice [4–6]. This law suggests that learning does not occur at a constant rate – when learning a new task, speed of performance's improvement declines. This variance of improvement should also affect the recognition rate if this new 'skill' was used as a behavioural biometric feature [4]. Haasnoot *et al.* [4] discuss the effects of practice and time on the behavioural biometrics recognition performance, working with a dataset in which subjects performed a 6-element discrete sequence production (DSP) task [7] with each participant completing 864 trials. They investigate how usage of initial samples (when a subject starts learning to perform a DSP task) and time between enrol and probe sessions affects performance. The authors find that early samples negatively affect recognition performance – this reflects the 'power law of practice'. Even after a recognition plateau is reached there is evidence that behaviour patterns keep changing – this is in line with the known facts about motor sequence learning – e.g. *motor chunks* [8] for DSP task and other tasks connected with motor memory. Both these findings support the idea that it is crucial how data is acquired and selected for behavioural experiments. For example, if behavioural data is acquired with a device that does not belong to the subject or subject needs to perform a specific task; there should be allowed enough time for the user to get accustomed to the device and enough attempts to learn the new task.

The fact that behavioural biometrics can include the need for subjects to learn tasks and that humans can change their behaviour is why this research domain is challenging. A disadvantage of (continuous) behavioural authentication methods is that it can not cope with unusual behaviour which can be caused by alcohol or injuries [9].

Researchers nowadays even offer open-source, extensible behavioural biometrics framework for Android called Itus [10]. The framework enables real-time classification on resource-constrained mobile devices for prototyping and deployment of new behavioural biometric schemes. This framework is widely used, including [11–13].

It is hard to find a sensor that is embedded in a mobile/wearable device which can not be exploited to extract behavioural biometric data. Various types of mobile devices are used to capture sensor data, ranging from wearable sensors to tablet devices. This is why in this paper we want to investigate data in behavioural biometrics and how the behavioural data is used. We only examine papers, that introduce new datasets. If a paper introduces a dataset but does not perform any experiments or the baseline results (as in [14]), we look into the same author successive work that presents such results. Especially, we focus on papers that:

- promote reproducible research by offering public datasets or using publicly available data;
- test generalisation properties of developed algorithms by using different subject data for training and testing or by using multiple datasets.

The remainder of this paper is structured as follows: Section 2 presents related works – different surveys of papers published in the field of behavioural biometrics; in Section 3 we present the motivation for this work and specifics about biometric data collection. Section 4 presents what we are examining when looking at the papers that are introducing new datasets; Section 5 – what

the sensors currently used in behavioural biometrics are. The main contribution is in Section 6 presenting the summary of 40 articles. In Section 7 selected papers are analysed in detail. We discuss the General Data Protection Regulation (GDPR) and its importance to research in biometrics in Section 8. Conclusions are given in Section 9.

2 Related work

Multiple review papers summarise and categorise existing work in mobile behavioural biometrics.

Alzubaidi and Kalita [15] present an extensive study of current (paper was published in the year 2016) research about behavioural biometrics. Authors analyse more than 70 studies and use seven behavioural biometric feature categories: hand waving (2 studies); keystroke (9 studies); touch screen (22 studies); gait (13 studies); signature (11 studies); voice (5 studies); behaviour profiling (9 studies). The paper also presents lessons learned, open problems and future trends including the opinion that behavioural biometrics are considered promising for providing continuous authentication for consumers; machine learning algorithms are well-suited to generalise from past user behaviours to '*predict the future*'; most current studies gather and record data under laboratory conditions; and that most published methods have been tested on the Android platform and ignored other platforms, e.g. iOS and Windows Mobile.

Rybnicek *et al.* [16] present an overview of biometric traits that can be used to secure mobile devices. The authors describe in detail keyboard, touchscreen, accelerometer, gyroscope based as well as hybrid authentication methods. Moreover, the authors discuss multiple aspects of behavioural biometrics for mobile devices and offer a roadmap. These aspects are criteria for using biometric data; data collection; system architectural structure; and biometric features and classification.

In [17] behavioural biometric systems are referred to as transparent authentication systems. The paper presents a review of these systems for mobile device security. The authors classify behavioural biometric systems into six categories: keystroke based authentication; gait based authentication; touch-based authentication; device sensor-based authentication; behavioural profiling based authentication; and multi-modal transparent authentification and analyse 33 papers published between the year 2007 and 2015.

Meng *et al.* [18] summarise biometric authentication methods on mobile phones. In their study, the authors classify biometrics used in 11 groups. Five physiological: fingerprint, face, iris, retina, and hand/palm recognition and six behavioural: voice, signature, gait, behaviour profiling (defined by authors as 'techniques that aim to identify people based upon the way in which they interact with the services of their mobile devices'), keystroke dynamics, and touch dynamics. The authors also propose a framework for establishing a reliable authentication mechanism through implementing multimodal biometric user authentication. Only there seem to be some inaccuracies in the mathematics used when *expected average* FRR (False Recognition Rate) and FAR (False Acceptance Rate) is calculated. The authors claim that both FRR and FAR can be improved by five orders of magnitude when the user has given three authentication tries and an option to enter a PIN if the person has not been recognised during the biometric authentication stage (similarly as the authentication mechanism in iPhones with Touch ID). Although, this would improve the FRR (since the genuine user has more attempts to get accepted), but this is not true for the FAR, since more tries would suggest that there is a more significant probability of getting accepted incorrectly/guessing the PIN.

In [18], published in 2015, the authors survey eleven biometric modalities used in smartphones – five physiological and six behavioural. The behavioural biometric modalities, that can be implemented in a mobile device and are summarised in this paper are voice recognition; signature recognition, gait recognition, keystroke dynamics, touch dynamics and behavioural profiling (in this case, the hypothesis is that the mobile users use applications differently depending on the location and the time). The authors also describe a generic biometric authentication system and present eight possible attack points summarising possible [practical] attacks and possible countermeasures. They also present guidelines for developing a robust biometric system concluding that multi-level authentication is preferable.

This 2018 survey by Gupta *et al.* [19] as its title would suggest – Demystifying Authentication Concepts in Smartphones – serves more as an explanatory dictionary for the novices in the field of mobile biometrics. One of the authors' objectives is to help explain to the new researchers the sophisticated jargon of biometrics. Similarly, as in our paper, the authors observe that usually, the mobile biometric systems are being reported in terms of accuracy, while other aspects are being overlooked. The authors emphasise that usability is a vital aspect to investigate, whereas, in our paper, the focus is the data itself – how the dataset is collected and used. The paper summarises relevant, publicly available behavioural biometrics datasets and briefly introduce the reader with research papers presenting recognition systems using these behavioural biometric modalities – gait, keystroke/touch dynamics, voice.

Buriro *et al*. [20] (published in 2017) is not a typical literature summary paper – instead, authors in this six-page long article discuss mobile biometrics and present 17 guidelines for designing and testing such systems. However, the focus is on both kind of mobile biometric systems – physical and behavioural ones. Authors conclude that for maximisation a proposed biometric solution's benefit; it is essential to evaluate the solution by using multiple criteria.

2.1 Gap in state of art

We did not find existing reviews that specifically discuss the collection and usage of behavioural biometrics data. That is why we offer this literature overview, presenting a summary of 40 research papers' approach on the collection and use of behavioural biometric data. Except for [20], currently, we do not see publications that would present extensive guides into how behavioural biometric experiments should be performed. Because of the lack of guidelines, well-established recommendations and handbooks in the behavioural biometrics field for the new scientists it is hard to navigate in this broad subject. In the last decade, there are thousands of published research papers in this field. Due to limited resources, we can not summarise them all. Instead, we have selected 40 papers in which authors present new mobile behavioural biometric datasets and perform experiments on the collected data. We choose papers from the last two decades, especially focusing on recently published papers, selecting papers that explore different mobile sensors and underlying behavioural modalities. We hope to provide an insight into how different researcher teams approach data collection and use.

3 Data collection and use

Li and Jain [21] states that 'the heart of designing and conducting evaluations is the experimental protocol. The protocol states how an evaluation is to be conducted and how the results are to be computed'. Our investigation suggests, researchers do not disclose much about data collection or how the collected data are used – what are the protocols and percentages of data used for training and testing. Often the reader can conclude, that collected data is used in an *All VS All* scenario. Generally speaking, this means that all data was available to the researcher during the development of the recognition algorithms. One or some samples of each subject's data are used as templates and the rest of the samples as probes (see definition in Table 1). *All VS All* means that all each subject's probes were scored against every subject's template, e.g. if there are N subjects, each with k samples and for each subject there is 1 sample used as a template, then all remaining subject's samples $(k-1)$ are scored against all subject templates (N templates), generating $N^2 \cdot (k-1)$ scores, $N \cdot (k-1)$ genuine and $N \cdot (N-1) \cdot (k-1)$ zero effort impostor scores.

Often no generalisation (algorithm testing on multiple databases/data subsets with non-overlapping subject data) of developed algorithms is researched – it is not clear to the reader

Table 1 Vocabulary used in this paper, based on the ISO/IEC 2382-37:2017 [22]

Term	Explanation
biometric template (template)	[22], enrolment data, stored for reference.
Biometric probe (probe)	[22] incoming biometric sample that is compared to the stored template.
enrol dataset	Collection of biometric templates, e.g. in training dataset.
probe dataset	Collection of biometric probes, e.g. in training dataset.
test dataset	The dataset that is used for assessment of the generalisation error of the final chosen model. Ideally, the test set should be kept in a 'vault', and be brought out only at the end of the data analysis [23]. The test set consists of enrol templates and probes. Test set should be solely used to report error rates and performance curves [24, 25].
Training dataset	Dataset that is used to fit the models [23] and train the classifier.
Validation dataset	Dataset that is used to estimate prediction error for model selection [23], and to estimate system's thresholds. Validation set consists of enrol templates and probes.
closed-set identification	Opposite of 'Open-Set Identification' [26].
open-set identification	It is unknown whether the subject presented to the biometric system for recognition has been enrolled in the system or not. Therefore, the system needs to decide whether to reject or recognise him as one of the enrolled subjects [26].
attempt	Submission of one (or a sequence of) biometric samples to the system [27].
identification	The process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual [22]. Also referred to as 1:*N* matching [21].
session	Data collection process separated by at least one day.
verification	The process of confirming a biometric claim through biometric comparison [22]. Also referred to as 1:1 matching [21].

Table 2 Division of a database's data into subsets (as used in [24, 25])

Subset	Subcategories	Purpose
train	—	train feature extractor and classifier
validation	templates; probes	estimate thresholds
testing	templates; probes	report recognition performance

how the developed system will perform with unseen subject data or if the system was over-fitted. Thus, it becomes hard to evaluate both – the modalities used and the developed algorithms.

The authors of this paper think that data collection, preparation (design of data protocols, the division of data into different datasets), as well as testing how the developed algorithms would generalise across unseen subjects are some of the most critical stages in the whole process of biometric research. Unfortunately, often in the literature, there is limited information on how this process should be done. In books about biometrics, there can be no information – e.g. [28] or description of basic – *All VS All* protocol – e.g. in [29].

In the literature about machine learning, especially deep learning, the importance of data and how data should be *treated*, as well as testing of algorithm generalisation is much more widely discussed. It is well-known in the machine learning community that it is vital to have separate training, test, and validation sets [23, 30]. 'The training set is used to fit the models; the validation set is used to estimate prediction error for model selection and the test set is used for assessment of the generalisation error of the final chosen model. The test set should be kept in a *vault*, and be brought out only at the end of the data analysis' – [23]. A typical split might be 50% for *the* training dataset, and 25% each for *the* validation and testing datasets. The validation set (but not the test set) can also be approximated by data re-use, by using cross-validation and bootstrap methods [23]. Luckily, some papers do discuss data importance for biometric experiments. [31] claims that to carry realistic (and unbiased) experiments, it is necessary to use different populations and data sets for development and evaluation. [32] investigate the effect on the performance of violating the rules for creating training sets of 'the Good, the Bad, and the Ugly' (GBU) biometric recognition challenge problem. The authors show that disregarding the GBU protocol can substantially overestimate performance on the specific face recognition task. Verification performance on the most challenging dataset – *ugly* increases from 11.4% (no subject overlap) to 15% when 91 subjects (out of 222 subjects) overlap in training and testing data sets (same subject different images). If images are drawn directly from the test set, verification performance increases hugely to 61.2%. The authors conclude that 'there are applications of face recognition where training and testing on the same people may be a reasonable thing to do, such as with family photo libraries'. They also recommend that researchers publicly post their training sets. This should provide confidence in the veracity of the reported results.

Other researchers, as suggested similarly in the machine learning literature, introduce three subsets – training, development (in other literature and this paper *development* data set is called *verification*) and testing. There are no overlapping subjects between these subsets [24]. The authors claim that 'such choice guarantees that specific behaviour (such as eye-blinking patterns or head-poses) are not picked up by detectors and final systems generalise well'. The authors recommend that training and development samples be used to teach classifiers how to discriminate. The training dataset can be used for training the classifier and the development data to estimate when the training should be stopped. Another way, which may generalise less well, is to merge both training and development datasets; use the merged set as training data and to formulate stop criteria. Finally, the test set should be solely used to report error rates and performance characteristics. Such, database division into three subsets and purpose are summarised in Table 2. If a single number is desired, a threshold τ should be chosen at the development (verification) set, and the half-total error rate reported using the test set data [24, 25].

The factors that we find essential when collecting and managing data – a roadmap to data use – are summarised in Table 3.

4 Assessment

In this section, we explain to the reader how we are assessing the 40 papers that are viewed in this survey. The aim of this paper is not to analyse all articles that introduce new mobile behavioural biometric datasets. Instead, we try to give the reader an insight into how data are being collected and used in this field.

In most of the papers surveyed in Table 4 data is collected so that researchers could develop the recognition algorithms and perform biometric experiments, assess ing the performance of developed recognition system s. The exception is the 'MIT Reality Mining Dataset' [38]. This study explores the capabilities of the smartphones and enables social scientists to investigate human interactions beyond the traditional survey-based methodology, rather than as a specific study about behavioural biometrics. Another exception is [39], where researchers have a different approach. They choose a novel behavioural biometric modality – the time it takes for the user to slide his/her finger between the unlock pattern points. They select machine learning algorithms (one-class support vector machine (SVM) and *K*-means) that can distinguish between the genuine and imposter behavioural patterns.

Table 3 Roadmap – set of actions when collecting and using data

Action	Reason
Acquire explicit consent from subjects; implement necessary technical and organizational measures for data collection, storage and usage.	General Data Protection Regulation [33].
Seamless data collection using subject's device OR (allowing subject to practice/learn task AND allowing subject to get used to the device).	'Power law of practice' [4–6].
Allow time between template and probe dataset collection.	To explore behaviour pattern changing, e.g. due to motor sequence learning e.g. because of *motor chunks* [8]; Saeed [34] discusses necessity to acquire data over multiple days in at least three sessions; ISO/IEC 19795-1 [27] states that enrolment and testing is normally carried out in different sessions, separated by days, weeks, months or years, depending on the target application.
For verification experiments – (false accept rate (FAR) AND false reject rate (FRR)) OR ROC OR DET curves should be reported.	ISO standard [35]; the best way to present or compare biometric verification performance is the ROC curve [36].
For open-set identification experiments – (False positive identification rate AND corresponding false negative identification rate) OR (False match rate AND false non-match rate) should be used.	ISO standard [35] (Section 6.3.4 – Identification metrics); [21].
Using training, testing and validation data sets.	Feature space and the verification system parameters must be trained using completely independent data from that used for specifying client models [31].
Assessment of generalization by testing algorithms using multiple, publicly available databases.	To see how good algorithms generalise on different datasets, and compare results with other researchers.
Results should be reproducible/reproducible research approach should be used.	To show evidence of the correctness of one's results and to aspire to reproducibility is to enable others to explore methods used and results acquired [37].

After selecting the algorithm and developing an Android app, the authors measure the system's recognition performance. In this case – not by creating a database. Instead, the authors recruited 54 subjects who act as genuine users and 10 subjects who act as imposters. In this way, observing the participants and counting, how often they succeed authenticating in the app, they determine the performance.

4.1 Data used

In the papers presenting new datasets, we examine what sensor data is collected. Due to space constraints, we describe only the type of sensors used, not specific usage details (e.g. the particular touch events are examined). The main data categories used are summarised in Section 5. We have summarised the subject count, whether subjects needed to perform a specific task (everything from walking [61] to navigating maps [14]). Because it takes time to learn a task and that subject's behavioural patterns can change over time, we examine whether researchers collect multiple session data, and if so, the time gap between sessions. It is also possible to have an 'unconstrained' data collection, when data are captured over multiple days and does not require the subject to interact with the system in any particular way, nor perform any task. We are interested in the overall duration of the data collection (sum of all data collection session duration).

In some studies, it is argued that subjects should not know the real data capture reason in order not to affect the way how subjects interact with the mobile devices during data collection, e.g. [13]. Of course, it may be perceived to be unethical, if an 'unconstrained' data collection's subjects would not know the collection reason and the type of data that is collected.

Most of all we think that it is essential if authors publish data used in experiments and provide others with the data protocols so that others can reproduce their experiments identically.

4.2 Experiments performed

We are not examining performance accomplished by the algorithms used, since a system's performance is enormously affected by the data used. For example, the number of subjects in the database, the data usage protocols (the instructions of which probes should be scored with which templates), feature/score fusion used, quality of the data (e.g. is the data collection unbiased, were all participant data collected in the same way). Some datasets are more challenging, allowing subjects to use the mobile device freely and consisting of larger number of subjects. It would not mean that algorithms developed using this dataset are *worse*, though the performance on this challenging dataset is lower. This is why we are investigating only general facts that would give an insight of how data was treated while performing experiments. We will ask whether authors explore generalisation (is the classifier is designed to correctly classify unseen objects which are not used during the training process? Generalisation represents the capacity of the classifier to respond to this task. When a classifier has a good generalisation capacity, it can correctly classify unseen examples [79]. To investigate, whether generalisation is performed, we are going to separately evaluate whether authors train system parameters and feature space on independent data (*Multiple datasets used*) or use multiple datasets (*Multiple datasets used*)). We also enquire, if open-set identification is performed, whether new subjects can easily be enrolled into the database, and what kind of data was used as imposters. An extended summary of parameters evaluated can be seen in Table 5.

4.3 Vocabulary used

There is a frequent confusion with terms used in behavioural biometrics studies even to the point of consistency of the nomenclature of behavioural biometrics. The names used in different studies:

- continuous authentication – [13, 14, 52, 53, 56, 71];
- continuous and passive authentication – [57];
- active authentication – [71];
- non-intrusive authentication – [54];
- non-intrusive user verification – [45];
- transparent authentication systems – [61];
- unobtrusive authentication – [61];
- implicit user identity recognition – [51];
- implicit authentication – [10, 41].

On account of this, we summarise vocabulary used in this paper in Table 1 based on the ISO/IEC 2382-37:2017 standard [22]. For

example, some papers, e.g. [67] use terms 'training' as building an enrol model and 'verification' as probing templates against probes, e.g. [52]) 'training' means enrol templates and 'testing' – to score them against probes. We will use terms as defined in the machine learning community, as identified in, for example, [24, 25].

5 Sensors used in behavioural biometrics

In mobile behavioural biometric studies, data are collected using mobile devices – phones, tablets and wearable devices. In this paper, we will not be looking at traditional biometric modalities

Table 4 Summary of the viewed research papers. Legend: □ – other; ∅ – not mentioned; ◐ – condition is partially true; ○ – condition is false; ● – condition is true

Author	Cit. per year	Year	OS	Data published	Software published	Sensors used	No. of subjects	Purpose known	Unsupervised	Data collection conditions	Specific task	No. of sessions	Days between sessions	Duration, approx	Multiple datasets used	Multiple databases used	Open-set identification	New subjects easily enrollable	Impostors	Comments
Eagle et.al. [38]	206.57	2006	Symbian		○	call logs, location, app usage, phone status	100	∅		various	○	1	-	9 months	-	-		-	-	MIT Reality Mining Dataset, discussed in detail.
Shi et.al. [40]	29.78	2011	Android	○	○	SMS meta-data, call meta-data, browser history, location, contextual data	276(50)			various	○	1	-	12 days	○	○		○	others and synthetic data	
De Luca et.al. [41]	55.75	2012	Android		○	touch	34	○		various		21	0	1.5 hours	○	○	○		others	Discussed in detail.
Feng et.al. [42]	31.75	2012	Android + wearable device	○	○	touch, accelerometer, gyroscope	40		○	lab		1	-	30 min	○	○	○	○	others	
Frank et.al. [13]	66.38	2012	Android		◐	touch	41	○	○	lab		2	7	60 min	○	○	○	○	others	Discussed in detail.
Kolly et.al. [43]	5.63	2012	Android	○	○	touch	14890	∅		various		-	-	∅	∅	○	○	○	others	Recognition performed for small subject groups - 2 to 15 people.
Sae-Bae et.al. [44]	35.75	2012	iOS	○	○	touch	34		○	lab		1	-	30 min	○	○	○		others	
Zheng et.al. [45]	27	2012	Android	○	○	touch, accelerometer	80		○	lab		1	-	5 min	○	○	○		others	
Damopoulos et.al. [46]	7.71	2013	iOS	○	○	touch	18			various	○	1	-	24 hours	○	○	○	○	others	
Li et.al. [47]	27.71	2013	Android	○	○	touch	75	∅		lab	○	1	-	20 min	○	○		○	others	
Zhao et.al. [48]	8.86	2013	Android	○	○	touch	30		○	lab		6	3	1 hour	○	○	○		others	
Antal et.al. [49], [50]	1.17, 3.33	2014	Android		○	touch	71	∅	○	lab		4	7	1 hour	○*	○	∅	○	others	*2 data subsets contracted for recognition, but the parameters tuned for each separate. Discussed in detail.
Feng et.al. [51]	15.83	2014	Android	○	○	touch, app usage	123(23)			various	○	1	-	21 day	○	○			others	
Yang et.al. [14], Sitová et.al. [52]	2, 33.75	2014, 2016	Android		○	accelerometer, gyroscope, magnetometer, touch	100 (90)	∅	○	lab		1	-	4 hours	○	○	○		others	Discussed in detail.
Khare et.al.[53]	0	2015	iOS	○	○	touch	20	○	○	∅		1	-	60 min	○	○	□	○	others	
Lin et.al. [54]	0	2015	Android	○	○	touch, accelerometer, magnetometer	35	○	○	lab	∅	1	-	3.5 hours	○	○	○	○	others	
Neal et.al. [55]	2.6	2015	∅	○	○	app usage, location	200	∅	○	various	○	1	-	19 months	○	○	○	○	others	
Roy et.al. [56]	2.6	2015	Android	○	○	touch, accelerometer, gyroscope	42	○	○	lab		1	-	30 min	○	○	○		others	
Wu et.al. [57]	1.6	2015	Android	○	○	touch	10	∅	∅	lab	∅	1	-	60 min	○	○	○	○	others	
Yang et.al. [58]	8.2	2015	Android (wearable device)	○	○	accelerometer, gyroscope	30	∅	○	lab		3	5	1 hour	○	○	○		others	
Abate et.al. [59], [60]	1, 6	2016, 2017	Android		○	accelerometer, gyroscope	100		○	lab		3	4	30 min	○	○	○		others	Discussed in detail.

Al-Naffakh et.al. [61]	3	2016	Android (wearable device)	○	○	accelerometer, gyroscope	10	∅	○	lab		2	∅	10 min	○	○	○	others	
Antal et.al. [62]	2.5	2016	Android		○	touch, accelerometer	54	○	∅	∅		3	7	3 hours	○	○	○	others	Discussed in detail.
Inguanez et.al. [63]	0.5	2016	Android	○	○	touch	32		○	lab		1	-	15 min	○	○	○	others	Only 1 subject was enrolled, rest treated as impostors.
Kumar et.al. [64]	5.75	2016	Android	○	○	touch, accelerometer	28		○	lab		7	0	2.5 hours	○	○	○	others	
Phillips et.al [65]	0.25	2016	iOS	○	○	location, accelerometer, gyroscope, magnetometer, phone status	4		○	lab		1	-	2.5 min	○	○	○	others	
Temper et.al. [9]	0	2016	Android	○	○	touch	25			various		1	-	2 weeks	○	○	○	others	
Anjomshoa et.al. [66], [67]	0.67, 11.33	2017	Android	◐	○	app usage, location, data usage	6	∅		various	○	1	-	76 days	○	○	○	others	
Feng et.al. [68]	0.67	2017	Android	○	○	call data, app usage, browser history, phone status, location	5	∅		various	○	1	-	50 days	○	○	○	others	
Fridman et.al. [69]	33.33	2017	Android	○	○	stylometry, app usage, browser history, location	200			various	○	1	-	30 days	○	○	○	others	
Papavasileiou et.al. [70]	1.33	2017	Other (Smart Socks)	○	○	textile pressure sensors, accelerometer	8		○	lab		1	-	5 min	○	○	○	others	
Shen et.al. [71]	15	2017	Android	○	○	touch, accelerometer, gyroscope magnetometer	102 (89)	○	○	indoor	○	3	2	63 hours	○		○	others	Algorithms tested on 2 datasets, not mentioned if parameters tuned.
Smith-Creasey et.al. [72]	1.67	2017	Android	○	○	accelerometer, gyroscope, magnetometer, proximity, ambient lighting, gravity, pressure sensor, location, touch, user activity	6	∅		various	○	1	-	14 days	○	○*	○	others	*2 datasets merged into one.
Buriro et.al. [73]	0.5	2018	Android	○	○	accelerometer, gyroscope, magnetometer, gravity, touch	85		○	lab		3	0	1 hour	○	○	○	others	
Li et.al. [74]	1	2018	∅	○	○	app usage, accelerometer, gyroscope	304	∅	∅	∅	○	1	-	∅		○	○	others	Data divided in *dev* and *eval* subsets.
Teh et.al. [75]	6	2018	Android		○	touch	150	∅	∅	various		1	-	20 min	○	○		others	50 and 150 subject subsets with overlapping subjects. Discussed in detail.
Leyfer et.al. [76]	0	2019	Android	○	○	touch	10			various	○	1	-	14 days	○	○	○	others	
Meng et.al. [77]	0	2019	Android	○	○	touch	50/16			various	○	1 / 2	- / 14	2.5h/3h	○	○	○	others	Only 16 subjects participated in the 2nd session.
Torres et.al. [39]	0	2019	Android	○	○	touch	54		○	lab		1	-	30 min	○	○	○	recruited subjects	Authors develop app first, then acquire results in a real-life scenario.
Yang et.al. [78]	1	2019	Android	○	○	touch	45	∅	○	various	○	1	-	14 days	○	○	○	others	

used in experiments, as undertaken in some papers. For example, Samangouei *et al*. [85] use face images as the only biometric modality, but Shi *et al*. [86] use voice as one of the modalities. Almost all data are in the form of time sequences (e.g. accelerometer data are acquired over time with a specific sample-rate). The time dimension will not be separately discussed. Each sensor data is described in detail, giving examples if possible. For

Table 5 Explanations to the parameters used in Table 4

Term	Explanation
Author	The first author of the paper. More than one paper can be mentioned, e.g. if dataset and the baseline experiments are presented in different papers.
Cit. per year.	Paper's citations per year. To calculate this parameter Google Scholar (https://scholar.google.com/) data was used. Data were acquired on 11.09.2019. Citations per year C_{py} is calculated as follows: $C_{\text{py}} = \frac{C}{2020 - y}$, where C – all paper's citations, y – the year in which paper was published. Because this survey does not include papers published after the year 2019, there cannot be the division by zero.
Year	Year in which paper (papers) were published.
OS	OS used on the mobile device for data collection.
Data published	Is data publicly available? (yes/no)
Software published	Is software used to perform experiments and produce results published? (yes/no)
Sensors used	List of sensors which data is collected. For the explanation of different sensors see Section 5.
No. of subjects	The number of subjects. This can be given as $N(K)$, where $N > K$. In this case, N is a number of subjects in the used dataset, but K – the number of subjects used in experiments. This would mean that part of subjects were excluded from experiments. In some open-set identification papers K is the number of enrol templates, and N subject data are used as imposters.
Purpose known	Do the subjects know the purpose of the data collection at the time when data collection takes place? (yes/no)
Unsupervised	If data collection were unsupervised? (yes/no) Supervised data collection would mean that at all times there was a person who interacts with the subjects (biometric attendant [22]).
Specific task	Did users perform a specific task? (yes/no) The opposite would be, that users were asked to use their mobile devices as usual, and data were collected in the background. The specific task cannot be related to the actual mobile device's use (e.g. in [61] the task is to walk).
No. of sessions	The number of data collection sessions. Although in some papers different definitions are used, we define a session as data collection process separated by at least one day. This means that 2 data collection attempts in a single day are counted as single data collection *session*. Continuous data collection, which spans over multiple days we define as a single session.
Days between sessions	Days between sessions. If there are multiple sessions with irregular time gaps between them, an average amount is given. If data collection was performed on consecutive days, 'days between sessions' = 0.
Duration, approx	Duration of the **entire** data collection (the sum of individual attempt and session duration).
Multiple datasets used	Were multiple datasets used (e.g. one for training feature space and fixing thresholds, other to obtain recognition results) with non-overlapping subjects? (yes/no)
Multiple databases used	Were multiple databases used to research generalisation? (yes/no)
Open-set identification	Was the developed recognition algorithms tested on open-set probe set? (yes/no) It is unknown whether the subject presented to the biometric system for recognition has enrolled in the system or not. Therefore, the system needs to decide whether to reject or recognise person as one of the enrolled subjects. It is the opposite of 'Closed-Set Identification' [26]. We will use this definition, similarly used in [80–84]. Li and Jain [21] define two metrics for open-set identification: False Alarm Rate and Detection Identification Rate. The Detection Identification Rate is the fraction of the genuine probes (that belongs to an enrolled subject) that are correctly detected and identified. The Detection Identification Rate is a function of the operational threshold. The Detection Identification Rate can be determined at a specific rank n, this requires that after matching, all similarity scores between genuine probes and the enrolled client templates are examined and sorted. A probe has rank n if the similarity score between it and its true template in the gallery is the nth largest similarity score. False Alarm Rate provides performance when a probe does not belong to any subject enrolled in the database [21]. We will not be so formal; only assessing what the probes are (e.g. are present probes of subjects, that are not enrolled in the template subset, so-called unknown-unknowns [81, 82]), not what are the metrics used.
New subjects easily enrollable	Can new subjects be easy enrolled? (yes/no) Whether it is easy to enrol new subject is a qualitative measure, but the authors of this paper think that it is vital to be assessed. The new subject enrolment process is straightforward with most hand-crafted algorithms, simple similarity metrics, hidden markov model (HMM), and if machine learning is used and the problem is approached as a one class problem (the one class is composed of the genuine templates, the classifier is unaware of the imposter data). New subject enrolment is difficult if a classifier with k output classes, where k – count of enrolled subjects, is used. We also argue that if the recognition is addressed as a two-class problem where one class represents the genuine templates, and the other class – imposters, it is difficult to enrol new subjects, since data to build both class models are available before a subject is enrolled. We believe that a more convincing method should be used. Zheng *et al*. [45] go further and say that 'two-class classifier need input data from impostors or non-target users at the training phase, which is unrealistic and raises privacy concerns'. Lui *et al*. [32] states that 'there are applications of face recognition where training and testing on the same people may be a reasonable thing to do, such as family photo libraries. However, it is entirely unacceptable for large scale deployed systems that must manage many enrolled people. For example, retraining a deployed system each time that a new person is enrolled is a logistical nightmare.'
Imposters	What are the imposter probe samples, when a system is evaluated. For example other subject data, synthetic data.
Comments	Our comments about the paper.

readability purpose, we divide sensors into physical and software sensors.

Hardware sensors:

- *Accelerometer* – used in [14, 56, 61, 71] sensor that measures the acceleration of the mobile device [39]. Acceleration is measured in three dimensions – *X*, *Y* and *Z* (relative to the phone's orientation), meaning, that the sensor provides information about phone's movement. It is one of the two sensors that determines the position of a device [87]. Gravity data (the applied force of Earth's gravity (m/s^2) on the mobile device) can be calculated using accelerometer data together with a magnetometer and the gyroscope. A device's orientation as used in [54, 71] can be computed using accelerometer and magnetometer [88]. This is why from now on the use of *orientation* sensor data will be denoted as the use of *accelerometer* and *magnetometer*. The accelerometer can be used to extract information about a person's gait as well as other movements that are *transferred* to the phone.
- Shen *et al*. [71] focus on motion sensor output while subjects are performing touch-tapping and single-touch-sliding actions. When users touch a smartphone, the accelerometer measures acceleration force applied to it; the gyroscope measures the rate of rotation, magnetometer measures the ambient geomagnetic field for three axes. It is possible to also determine degrees of phone's rotation around three physical axes. Users may develop personal operational habits, which are based on a different rhythm, strength, and angle preferences of finger movements.
- *Geomagnetic field sensor* – monitors changes in the earth's magnetic field [89].
- *Gyroscope* – measures the rotation around a device's axis [90]. Accelerometer and gyroscope sensors are always hardware-based, and a variety of software-based sensors can use their data [91].
- *Location* – in this paper, denotes location through Wi-Fi information, cell tower information, global positioning system (GPS) and Bluetooth information.

One of the key ideas in [38] is to exploit the fact that modern phones use multiple networks (e.g. Bluetooth and global system for mobile communications (GSM)) data about multiple networks can complement each other to determine subjects' location and actions. Both radio tower IDs and nearby Bluetooth devices can be logged to obtain different types of information. Every Bluetooth device is capable of 'device-discovery,' which allows them to collect information on other Bluetooth devices within 5–10 m. This information includes the Bluetooth MAC address (BTID), device name, and device type. The BTID is a 12-digit hex number unique to the particular device.

The authors of [40] recorded GPS coordinates if users enabled it. They also collected location, obfuscated service set identifier (SSID) if a user is connected to a Wi-Fi network and IDs of cellular base stations.

In [69] location was determined using GPS data when outdoors and Wi-Fi when indoors.

- *Magnetometer* – provides general rotational information and relative orientation of the mobile device to the Earth's magnetic north [92]. The magnetometer is enclosed in an embedded device that often incorporates accelerometer. This helps correct magnetometer sensor measurements using tilt information from the auxiliary sensor. The magnetometer is similar to the geomagnetic field sensor, except that no hard iron calibration is applied to the magnetic field [87].
- *Phone status* – meta-data about the subject's phone.
- In [38] the authors collect information about phone status such as charging and idle. In [65] information about battery level was recorded and analysed.
- *Proximity sensor* – determine how close the face of a device is to an object [87].
- *Touch* – denotes the usage of the touchscreen.

In [14] *touch* describes raw touch events: a timestamp, finger count, finger ID, raw touch type, *X*/*Y* coordinates, contact size, screen orientation data, tap gesture data, scale gesture data, scroll gesture data, fling gesture data, key press on virtual keyboard data including press type, and key ID.

The authors of [13] define two user touch actions that are frequent and primitive and call them 'trigger-actions'. Whenever the user performs such action, the system records touch data. These actions are sliding horizontally over the screen (e.g. when browsing through images or navigating to the next page of icons in the main screen) and sliding vertically over the screen (e.g. to move screen content up or down, this is typically done for reading e-mail, documents or web-pages).

In [71] it was found that more than 98% of touch-interaction behaviour comprises touch-tapping and single-touch-sliding actions.

Software sensors:

- *App usage* – monitoring of applications being used.
- The authors of [66, 67] collect data from five popular social network services – Facebook, Twitter, LinkedIn, Skype and WhatsApp. Collected data includes a session ID, application name, the duration of that session, the amount of data used in the session and the initial location where the session started.
- In [69] the application's name is used as the unique identifier, and the number of times a user uses each application is counted. This means that application usage is not constrained to a number of applications. For each user, a classification model is constructed by determining the top 20 entities used by that user in the training set. Entities that do not appear in the top 20 are considered outliers and ignored.
- *Browser history* – monitoring of websites being visited.
- Browser history and obfuscated domain name of each URL visited was recorded by Shi *et al*. [40].
- Fridman *et al*. [69] used the domain of the URL as the unique identifier and counted the number of times a user visits each domain in the training set. For example, 'm.facebook.com' is a considered a different domain than 'www.facebook.com' because the sub-domain is different. Similarly, to application usage, a user's web history model was determined by the top 20 entities visited. Entities that do not appear in the top 20 are considered outliers and are ignored in this classifier.
- *Call data* – As *call data* we denote SMS meta-data, call meta-data and call logs.
- Eagle and Pentland [38] logged calls and SMS, call duration, unique callee identifier (natural number).
- In [40] the authors record the time and direction (incoming or outgoing) for SMS messages and calls. Data about call duration are collected. The authors did not collect subjects phone contact list and used obfuscate phone numbers.
- *Contextual data* – the study presented in [40] uses contextual information, such as the content of user's calendar entries.
- *Data usage* – monitoring of data being uploaded or downloaded.
- *Stylometry* – study of linguistic style. It has been extensively applied to the problems of authorship attribution, identification, and verification [69].

The authors in [69] collected text entered by subjects via a soft keyboard. The activity in the majority of cases was communication via SMS, MMS, WhatsApp, Facebook, Google Hangouts, and other chat applications. Text input was mostly completed in short bursts. A tracking application captured the keys that were pressed on the keyboard and not the auto-corrected result. The majority of the typed messages, therefore, had a considerable amount of misspellings and words that were erased in the final submitted message. In such a way, it was possible to acquire unique typing idiosyncrasies that auto-correct can conceal.

6 Discussion

In this section, where we present our chosen research papers that introduce behavioural biometrics datasets to investigate what the

behavioural biometrics trends are. Table 5 summarises the criteria used to evaluate papers, whilst Table 4 presents detailed information about 40 papers.

6.1 Data publishing effect on citation count

Analysing data summarised in Table 4 (if multiple papers were used to present one data set, they both were separately included; papers in which data were partially published (marked with ◐ in Table 4) were not included) we can conclude that the mean value of citations, if a paper presents a published data set, is 34.9. If a paper presents a data set, that is not made publicly available, the mean citation count is 8.7.

Sadly, there are too few studies that publish data (8 studies, 11 papers involved). Moreover, because of that, it is difficult to make any conclusions. It can be noticed that if data is published, it is possible for the paper to become more popular. As it can be seen with Eagle and Pentland [38] (MIT Reality Mining Dataset) with 206.6 citations on average per year, Frank *et al.* [13] with 66.4 citations per year, De Luca *et al.* [41] with 55.8 citations per year and Sitová *et al.* [52] with 33.75 citations per year.

However, there are widely cited studies, that do not publish data (Sae-Bae *et al.* [44] with 35.8 citations per year; Fridman *et al.* [69] – 33.3 citations per year; Feng *et al.* [42] – 31.8 citations per year; Shi *et al.* [40] – 29.8 citations per year; Li *et al.* [47] – 27.7 citations per year). These studies appear as outliers between papers that are not publishing collected behavioural data.

Looking at the ten most cited studies, we did not find any statistically significant evidence that a particular experiment setup – constrained versus unconstrained; asking the participants to perform a specific task versus using a device freely.

6.2 Trends in use of mobile operating systems

In most of the 40 reviewed papers, devices running the Android operating system are used. These mobile devices include wearable devices – Al-Naffakh *et al*. [61] use Microsoft Band 2, and Yang *et al*. [58] use the Samsung Galaxy Gear smartwatch. Use of the *Android* operating system is understandable since it gives higher freedom for the app developers than iOS where some concepts are difficult to be accomplished [53]. The 'MIT Reality Mining Dataset' [38] uses Nokia 6600 phones using Symbian Series 60 operating system (data were collected in the year 2004); the authors of [53, 65] use iOS devices (iPhone 5S and iPhone 6 accordingly). In [46], the authors recruited 18 subjects to collect data using their iPhones.

Two studies used a different kind of wearable devices: the authors in [70] used Sensoria smart socks that are embedded with three proprietary textile pressure sensors and a 3-axis accelerometer. The authors of [42] in addition to an Android device also use a digital sensor glove equipped with accelerometer and gyroscope sensor boards based on the Arduino platform. It provides the finger movement information. The authors collected two datasets – only touch data from the mobile device and touch data together with the finger movement data. The proved that the additional finger movement information improves the recognition rate.

6.3 Modalities used and subject count in presented databases

It can be summarised that different researchers' approaches to data collection are diverse: the duration of the data collection can span from ∼2.5 min as in [65] to 19 months (in 'Mobile Device Application, Bluetooth, and Wi-Fi Usage Data as Behavioural Biometric Traits' – [55]). The same applies to the number of subjects, ranging from 4 participants in [65] to 14,890 subjects in [43]. There are multiple notable examples, with the subject count over 100. For example, Li and Bours [74] with 304 subjects, Neal *et al*. [55] and Fridman *et al*. [69] with 200 subjects. Publicly available datasets with more than 100 subjects data are Teh *et al.* [75] dataset, published in the year 2018, containing 150 subjects data, 'MIT Reality Mining Dataset' [38], Hand Movement, Orientation, and Grasp (HMOG) dataset [14, 60], as well as a dataset published by Abate *et al.* [59, 60] – each with 100 subjects data.

How data are collected also differs considerably. Two extremes are entirely constrained, lab-based data collection and fully unconstrained when the mobile devices are given to subjects to use them as one's own. There are *in-between* studies: authors of [9] asked participants to use a self-made banking app. A similar approach is used by Kolly *et al.* [43] and De Luca *et al.* [41]. In [43] the authors developed a mobile game that was used by 14,890 subjects. In [41] the authors asked the subjects to input a password pattern that consists of five strokes. There were 21 data collection sessions. Thus, subjects were not limited to their choice of location/environment, but they still had to perform a specific task.

Talking about subjects – publication [78] is impressive with the fact that in their data collection, participated users as young as ten years old.

6.4 How were experiments performed?

A summary of the papers shows that in behavioural biometrics, recognition algorithms and thresholds are not trained using separate datasets. An exception is the authors of [74] – they use separate subject data for crafting of the recognition algorithms and testing the developed biometric system. Data collected from 304 subjects are divided into two datasets – development dataset containing 142 subject data for choosing the optimal weights, and evaluation dataset with 162 subject data on which to report the results. Another exception is in [43]. Its authors have collected data from 14,890 subjects and test the recognition performance on small subject groups – 2 to 15 people. The authors also claim that 'in all experiments, we have separated training and test datasets to avoid over-fitting', but, unfortunately, they do not provide any additional details. The authors of [71] tested algorithms on two datasets, but it was not disclosed whether the thresholds and other parameters were fixed for both datasets or fine-tuned for each.

In other papers, even if there were some division performed, all data subsets contained the same subject different data. Thus, the use of the data in behavioural biometrics is different from when dealing with traditional biometric modalities – e.g. facial recognition.

Across the papers, different biometric experiments are performed: verification, closed-set identification or open-set identification. In [53], verification experiments are conducted, and even the data were collected in a way that would represent verification (1:1 scenario) by asking imposters to enter the genuine user's user name.

In [40, 47, 51, 63] the authors use open-set identification. Shi *et al.* [40] developed a data collection application which was posted in the *Android Marketplace* and freely downloadable by anyone. Among the 276 users who downloaded data collection application, the authors selected 50 users (who participated over a period of 12 days or longer) as the genuine users and selected subjects who have participated over a duration of 3 days or longer to serve as impostors (so-called 'unknown-unknowns' [81]). The authors of [51] recruited 23 genuine users that use their own Android mobile devices and selected 100 impostors. A paper presented by Inguanez and Ahmadi [63] offers an interesting study – the authors analyse text typed using an on-screen keyboard, focusing on touch features and time intervals. They collected a dataset consisting of 32 subjects. Of all touch data, 48% is generated by one subject, who represents the genuine device user. The remaining 31 users represent the imposters. In [47] the authors selected 28 out of 75 subjects as *target users*. None of the papers offering open-set identification experiments used ranking, and it is understandable that this would not make much sense, since the task is to protect a single smartphone user, instead of finding similarity in a database. Open-set identification has the benefit of allowing the addition of subjects which have partially participated in the data collection. In this way, one can test the system more thoroughly, with a broader variety of imposters, since more imposters are scored against each genuine template. In the remainder of the papers, closed-set identification was performed.

Another aspect we were interested in is whether new subjects can be quickly enrolled in the system, in case such verification/identification systems are used in real-life scenarios.

We were pleased to find that in multiple cases, new subject enrolment is straightforward. Authors use simple similarity metrics – in [61] the authors use the average Euclidean distance between the reference template and probe sample; this distance value represents the similarity between both samples. Zhao *et al*. [48] employ as the similarity metric normalised cross-correlation – L_1 and L_2 norms. To compute similarity, Teh *et al*. [75] use a score level fusion using three simple similarity metrics – Gaussian Estimation, *Z*-Score and Standard Deviation Drift. The authors of [41, 58, 60] use Dynamic Time Warping (DTW). In [56], an HMM-based behavioural template training approach is presented. It does not require training data from other subjects other than the owner of the mobile device and can be updated with new data over time.

One-class classifiers are used, for example, in [52] where the authors performed experiments using three different one-class verifiers: scaled Manhattan (SM), scaled Euclidean (SE), and 1-class SVM, since using a one-class approach there is no need for imposter data during the training stage. Zheng *et al*. [45] use a one-class classifier, the nearest neighbour distance, to the training data. Yang *et al*. [78] found that the best performing algorithm to recognise whether a particular touch operation belongs to the genuine user is one-class SVM.

The paper by Kolly *et al.* [43] presents a behavioural biometrics dataset that is collected unusually. The authors developed a mobile game for the Android platform that was freely downloadable by everyone. The game's user interface used different elements that users had to touch: buttons, lists, and radio buttons. The mobile game was downloaded by 14,890 subjects. Overall, 1 million button touch events and over 2 million list touch events were generated while using the game app. Sadly, the authors do not present additional statistics about the collected data such as what was the average interaction time with the app and how long did the data collection continue. There also were not many details about how data was used. It was disclosed that subjects were not scored in an *All VS All* scenario. Instead, the recognition experiment was to recognise a person out of a set of 2 to 15 people, concluding that the recognition rate drops when increasing the number of subjects. The authors highlight and measure three properties of touch: timing, pressure, and the position relative to the target.

Neal *et al*. [55] present a prolonged study – authors have collected 200 subject data for 19 months. The data collected includes application data, Bluetooth and Wi-Fi information. The authors present recognition rates of four data types – application, Bluetooth, Wi-Fi and combined. They offer results for consecutive day, week, and month scenarios, concluding that there is a possibility that a day is too short a period in 'which to extract features' (at least for the used features).

7 Papers in detail

In this section, we will assess selected papers in detail. We have selected all eight papers that present public datasets.

- *Abate et al.* [59] – *'Smartphone Enabled Person Authentication Based on Ear Biometrics and Arm Gesture' and* [60] – *'I-Am: Implicitly Authenticate Me Person Authentication on Mobile Devices Through Ear Shape and Arm Gesture'*. The authors present a novel approach for incorporating physical and behavioural biometric features into a biometric authentication system. They use accelerometer and gyroscope readings to record arm gestures and use the phone's secondary (frontal) camera to acquire images of a subject's ear. The idea is to capture these modalities when the subject is answering a phone call; arm dynamics should affect the smartphone motion pattern due to behavioural and anatomical characteristics. Arm gesture data in conjunction with the ear biometrics makes a compelling and never-before viewed combination. The authors present a publicly available (upon request by e-mail) database, consisting of 100 subject arm gesture and ear data. 30 of the 100 subjects participated in 3 data collection sessions (spread over a two-week time period). The remaining 70 subjects participated only in the first session. Data is acquired using the Samsung Galaxy S4 smartphone.
- *Antal et al.* [49] – *'Identity information revealed from mobile touch gestures' and* [50] – *'Information revealed from scrolling interactions on mobile devices'*. The authors present a publicly available scrolling touch data database. It consists of 71 subjects' behavioural data. Data capture consisted of four sessions, each a week apart. Users were asked to perform two tasks – text reading that requires vertical scrolling and image selection, requiring horizontal scrolling. For each touch point timestamp, x and y coordinates, the pressure and the area under the finger were recorded. Antal *et al*. [50] published in 2015, extends Antal *et al*. [49] published in 2014. Antal *et al*. [50] not only presents the recognition experiment results but also tries to answer questions regarding obtaining information about the subjects themselves: can a subject's gender be determined using touch swipe data; what separates male and female touch data; is it possible to identify the subject's experience level of touchscreen usage.
- The authors incorporate subject data into four different datasets: Datasets 1 and 2 for subject classification, Dataset 3 for gender classification and Dataset 4 for touchscreen experience level classification. Unfortunately, there is no information given or how data is distributed across the datasets, e.g. why the gender classification dataset consists of just 18 subject data. Although there are two separate datasets for the subject classification task, the parameters for each dataset were fine-tuned separately.
- *Antal and Nemes* [62] – *'The MOBIKEY Keystroke Dynamics Password Database: Benchmark Results'* The work of Antal and Nemes examines how keystroke dynamics can complement passwords. For the data collection, the authors developed an Android app and used 13 identical Google Nexus 7 tablets. In the data collection participated 54 subjects. Data were collected in a single session, where participants were asked to enter pre-defined passwords (an easy one, logical-strong and strong). Each password had to be entered correctly at least 20 times – resulting in 60 samples per subject. The developed app gathered accelerometer and touch data. Interestingly, the authors present manually crafted so-called 'second-order' feature set where the number of features is independent of the password's length.
- *De Luca et al.* [41] – *'Touch me once and I know it's you! Implicit Authentication based on Touch Screen Patterns'*. The authors perform two sub-studies; the second study's database is made public (available upon request via e-mail). The authors pay great attention to the data collection process and the methods of how to make the study more unbiased. They investigate, how to equip a shape-based unlocking scheme with behavioural authentication methods. This is done by exploiting the touch data when the subject is entering the unlock shape.
- The first study (data are not publicly available) is a time-limited, lab-based, two-session study. Sessions were 2 days apart. In it, subjects were asked to unlock the data collection device by swipe (horizontal, vertical, vertical with two fingers and diagonal unlock swipe patterns). Subjects know the purpose of the data collection and are asked to unlock the device, in the same way, on every occasion. The authors gave the data collecting device (an Android Nexus One mobile phone) to the subjects in test mode, for them to get familiar with the task they needed to perform. In the actual data collection, authors used two devices – one for data collection and the other one for subject *distraction*. After every 20 unlock tries, subjects had to input text message's text in the other phone. The second session was undertaken to minimise the subjects' habituation effects.
- After this study, the authors performed a second experiment, reasoning, that not enough touch data was collected and the fact that the subjects know the original data collection's purpose. The follow-up study's data is publicly available. In this study, the authors investigate the possibility of adding additional protection to the password pattern by using behavioural biometric authentication. The password patterns consist of five strokes assigned randomly to the subjects. Subjects could

practise their issued password pattern as many times as necessary before the actual data collection. After the practising stage was completed, subjects were asked to execute one authentication per day for 21 days. Subjects could perform authentication in an unconstrained environment. After completing the 21 sessions, participants were asked to attend a follow-up event, where the data was copied from the devices and *imposter* data recorded. The premise was that the imposter knows the genuine user's password pattern. Since devices used by the participants varied, only the generated data of users, who owned the same type device were used as the imposter data.

- *Eagle and Pentland* [38] – *'Reality mining: sensing complex social systems'*. The authors introduce the 'MIT Reality Mining Dataset' that consists of 100 subject data collected over a period of 9 months. For the data collection, 100 Nokia 6600 mobile phones are used. Information about call logs, Bluetooth information, cell tower information, application usage, and phone status (such as charging and idle) are recorded. The dataset is publicly available. This study is undertaken to explore the capabilities of the smartphones and enable scientists to investigate human interactions beyond the traditional survey-based methodology, rather than as a specific study about behavioural biometrics. This data set was collected in 2004 and was first such large-scale database. Although initially intended to study social behavioural, researchers have used this dataset for behavioural biometric experiments, e.g. in [93].
- *Frank et al.* [13] – *'Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication'*. This paper's hypothesis is that continuously recorded touch data from a touchscreen is distinctive enough to serve as a behavioural biometric *modality*. Both horizontally sliding and vertically sliding touch actions are used. Android phones are given to 41 subjects; the authors try to use a minimal number of different phones (four different phone models, five devices in total) because using different mobile phones could artificially improve accuracy thus invalidating the results. During the data collection, subjects are unaware of the actual purpose of the data collection. Subjects read various articles (excerpts from Wikipedia articles) and find differences between cartoon drawings. There are two data collection sessions, a week apart. The authors divide data records into individual strokes and propose the use of 30 hand-picked features. Informativeness of these features and their mutual correlation are explored. Two different classifiers, *k*-nearest neighbour and support vector machines are used. First, a model is trained, after which the same model is tested with the same subject's unseen data. There are three application scenarios (inter-week authentication, intersession authentication and short-term authentication). Data and partial software (a *Matlab* implementation of feature extractor) is published.
- *Teh et al.* [75] – *'TDAS: a touch dynamics based multi-factor authentication solution for mobile devices'*. In this paper, the authors propose to implement the use of touch dynamics into a PIN-based authentication system. A Samsung Galaxy Tab 10.1 is made available for 150 subjects. They can choose their preferred location (e.g. home, work, a public space) and are allowed to familiarise themselves with the data collection application and procedure. The single-session data collection's duration is 15–20 min (not including the time it takes for the subjects to familiarise *themselves* with the task). The participants are asked to enter 4-digit and 16-digit PIN codes ten times each, resulting in 20 samples per subject. Authors have published three different versions of the dataset – subsets containing 50, 100 and 150 subjects. However, there are subjects whose data is present in all subsets, meaning that it is impossible to see how well the recognition algorithms work on unseen persons' data.
- *Yang et al.* [14] – *'A Multimodal Data Set for Evaluating Continuous Authentication Performance in Smartphones'* and *Sitová et al.* [52] – *'HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users'*. Yang *et al*. [14] introduce a comprehensive behavioural biometrics dataset. Data is acquired from 100 subjects, and there are 9 channels of data – accelerometer, gyroscope, magnetometer, raw touch event, a tap gesture, scale gesture, scroll gesture, fling gesture and key press on the virtual keyboard. Data were collected using an application developed by the authors for Android phones. Subjects were asked to perform three different type tasks: document reading, text production and navigation on a map to locate a destination. For each type of task, the volunteer either sits or walks while performing the task (thus there are two operational scenarios). One attempt lasts 5 to 15 min, and each volunteer is expected to perform 24 attempts (8 for each task type). In total, dataset consists of 2 to 6 h of each subject's behavioural biometric data. The dataset is publicly available.

Sitová *et al*. [52] offer baseline experiments for this dataset. The authors propose a new set of behavioural biometric features for continuous authentication, calling them 'Hand Movement, Orientation, and Grasp (HMOG)'. They propose the use of a one-class classifier [94]. From the original 100 subject dataset, 10 subjects are excluded due to an insufficient amount of data. To improve authentication performance, the authors performed feature selection, feature transformation with principal component analysis and outlier removal. For each verifier, HMOG features were selected separately. For this, the authors used 10-fold cross-validation.

Although the terms 'training' and 'testing' are used, in the paper they refer to enrolment template and probe data (e.g. 'we used all training vectors to construct the template'; 'we created authentication vectors by averaging test vectors'; 'authentication vector was matched against the template of the same user'). This suggests that an algorithm's ability to generalise to unseen subjects is not investigated (for example by dividing collected data into training, validation and test subsets, training the model on one subset and acquiring the recognition rates on another); the system is tuned for a particular set of subjects.

8 General data protection regulation

The GDPR [33] is a European Union (EU) legal requirement for data protection. GDPR enables citizens of the EU to control their data. It became enforceable on 25 May 2018 ([33], Article 99), after a two-year transition period, replacing the 1995 Data Protection Directive.

The GDPR Article 4.14 defines that biometric data is 'personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic (fingerprint) data'. This broad definition implicitly acknowledges that biometric technology will continue to evolve [95]. The definition recognises traditional biometric features as well as behavioural biometrics. The extent of the possible behavioural biometric features is not clear from this definition since the definition raises a question what does unique identification imply – is the purpose of the use of biometrics for identification/authentication enough to consider it to be a unique identification or will a certain level of precision have to be reached to assume that the biometric feature in question delivers *unique identification*?

The significant change regarding biometrics when the existing Data Protection Directive is compared with GDPR is that the GDPR recognises biometric data as a subset of sensitive personal data [95, 96]. The new directive, Article 9.1 states that processing of such data is prohibited. Fortunately, there are exceptions for processing 'special categories of personal data'. The general exception – if the 'data subject has given explicit consent to the processing of those personal data for one or more specified purposes' (Article 9.2a). This means that biometrics can still be used in various kinds of application but a user's consent is mandatory and a special care is necessary to collect, store and use the data. Use of special category data is still allowed for scientific research purposes (Article 9.2j).

Data processing is defined (Article 4.2) as 'any operation or set of operations which is performed on personal data or on sets of

personal data, whether by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'. The definition states that even **storage** of such data is something that falls under the GDPR and that requires special care; this also applies to the scientific community.

The most concern for the scientific community arises from the distinct care required with the special category data. GDPR Article 28.2 says that the processor (defined as 'natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller', GDPR Article 4.8) shall not engage another processor without prior specific or general written authorisation of the controller'. In other words, this mandates a strict system of how to access and use data inside scientific organisations including sharing the data across researchers and organisations.

GDPR Article 17 defines the Right To Erasure (or Right To Be Forgotten). The data subject, at any time, can withdraw his or her consent and their data must be deleted. GDPR Article 17.2. states that 'where the controller has made the personal data public and is obligated (…) to erase the personal data, the controller (…) shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.' This would imply that anyone at any time can ask to withdraw their biometric data from the database. If the data are published, the institution that published the data have to ask the third parties to delete the subject's data in question, but GDPR Article 17.2. releases the original authors from liability if the third party does not cooperate. The sharing of data still would mandate a written authorisation, as stated in Article 28.2.

GDPR Article 17.3.d states that Right To Be Forgotten does not apply to the extent that processing is necessary 'for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89.1', which comments on the necessary technical and organisational measures that need to be in place. Article 89.1 also states that 'Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.' The Right To Be Forgotten does not necessarily relate to the biometric databases if used for scientific purposes but it is not entirely apparent what the last sentence implies: does data subject pseudonymisation is enough or biometrical data should be distorted to not permit the identification?

GDPR Article 25 defines 'Data protection by design and by default'. The article states that the data controller shall 'implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects'. Similarly, as with the previous point, it is not apparent how extensive these measures should be.

Summarising the GDPR with respect to biometrics one can conclude that it tries to organise and regulate all the aspects of a natural person's data usage. It has an enormous effect on biometric data and its use in the scientific community. One positive outcome is that finally these aspects are being regulated at the *EU's* level. Unfortunately, there are currently many questions and no known precedent or guidelines how the necessary mechanisms (Right To Be Forgotten; Data protection by design and by default; data processing; data sharing) should be implemented and maintained while collecting, using and sharing biometric data for scientific reasons.

Irwin [97] states an interesting concept that entities like to experiment with new technologies (e.g. biometrics) because they are accessible. It is possible that because of the high data requirements set by the GDPR this trend will change, and companies will evaluate more carefully whenever they want to handle the special categories of personal data.

9 Conclusions

In [98] five levels of reproducibility are defined: Reviewable Research; Replicable Research; Conformable Research; Auditable Research; Open or Reproducible Research. Unfortunately, most of the papers published in the field of behavioural biometrics fall under the first level 'Reviewable Research' – 'The descriptions of the research methods can be independently assessed and the results judged credibly. This includes both traditional peer review and community review and does not necessarily imply reproducibility' [98].

Although we cannot prove that there is a correlation between data publishing and citation count, we still think that it is beneficial, if not for the researcher itself, then for the whole research community to publish databases and software for result reproduction. There is a gap for reproducible research and grand challenges to be made (similarly to the GBU [32] in face recognition) in the field of behavioural biometrics.

There seems to be a need for standardisation and guidelines on how biometric experiments should be performed to enable authors to implement how training is to be conducted and how to compute results.

The GDPR tries to regulate a somewhat grey area of personal data, including biometrics. Only time will tell how it will be interpreted by businesses, researchers and the general society. Regarding research in biometrics and GDPR – hopefully, issues will become more apparent and there will be precedents of implementation of the necessary technical and organisational measures.

10 Acknowledgments

Thanks to Erwin Haasnoot, Tiago Freitas Pereira, Sushil Bhattacharjee, Elakkiya Ellavarason and Matthew Boakes for the valuable discussions. We also want to thank the anonymous reviewers for their insightful comments.

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 675087.

11 References

[1] Bank of America: 'Trends in consumer mobility report 2016', 2016. Available at http://newsroom.bankofamerica.com/files/press_kit/additional/2016_BAC_Trends_in_Consumer_Mobility_Report.pdf, accessed 26 November 2017

[2] Gazzaley, A., Rosen, L.D.: '*The distracted mind: ancient brains in a high-tech world*' (MIT Press, UK, 2016). Available at https://books.google.co.uk/books?id=o8sbDQAAQBAJ

[3] Asarch, S.: 'South park' fort Collins: emoji analysis nails the future of communication', 27 October 2016. Available at http://www.player.one/south-park-fort-collins-emojianalysis-nails-future-communication-565459, accessed 25 November 2017

[4] Haasnoot, E., Barnhoorn, J.S., Spreeuwers, L.J., *et al.*: 'Towards understanding the effects of practice on behavioural biometric recognition performance'. 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018, pp. 558–562, DOI: 10.23919/EUSIPCO.2018.8553446

[5] Ericsson, K.A., Krampe, R.T., Tesch-Romer, C.: 'The role of deliberate practice in the acquisition of expert performance', *Psychol. Rev.*, 1993, **100**, pp. 363–406

[6] Anderson, J.R.: 'Acquisition of cognitive skill', *Psychol. Rev.*, 1982, **4**, pp. 369–406

[7] Abrahamse, E., Ruitenberg, M., De Kleine, E., *et al.*: 'Control of automated behavior: insights from the discrete sequence production task', *Front. Hum. Neurosci.*, 2013, **7**, p. 82. Available at https://www.frontiersin.org/article/10.3389/fnhum.2013.00082

[8] Acuna, D., Wymbs, N., Reynolds, C., *et al.*: 'Multifaceted aspects of chunking enable robust algorithms', *J. Neurophysiol.*, 2014, **112**, (8), pp. 1849–1856

[9] Temper, M., Tjoa, S.: 'The applicability of fuzzy rough classifier for continuous person authentication'. 2016 Int. Conf. on Software Security and Assurance (ICSSA), St. Pölten, Austria, 2016, pp. 17–23

[10] Khan, H., Atwater, A., Hengartner, U.: 'Itus: an implicit authentication framework for android'. Proc. of the 20th Annual Int. Conf. on Mobile Computing and Networking, Maui, HI, USA, 2014, pp. 507–518

[11] Bo, C., Zhang, L., Li, X.Y., *et al.*: 'Silentsense: silent user identification via touch and movement behavioral biometrics'. Proc. of the 19th Annual Int. Conf. on Mobile Computing & Networking, Miami, FL, USA, 2013, pp. 187–190

[12] Feng, T., Zhao, X., Carbunar, B., *et al.*: 'Continuous mobile authentication using virtual key typing biometrics'. 2013 12th IEEE Int. Conf. on Trust, Security and Privacy in Computing and Communications (TrustCom), Melbourne, Australia, 2013, pp. 1547–1552

[13] Frank, M., Biedert, R., Ma, E., *et al.*: 'Touchalytics: on the applicability of touchscreen input as a behavioral biometric for continuous authentication', *IEEE Trans. Inf. Forensics Sec.*, 2013, **8**, (1), pp. 136–148

[14] Yang, Q., Peng, G., Nguyen, D.T., *et al.*: 'A multimodal data set for evaluating continuous authentication performance in smartphones'. Proc. of the 12th ACM Conf. on Embedded Network Sensor Systems. SenSys '14, Memphis, TN, USA, 2014, pp. 358–359. Available at http://doi.acm.org/10.1145/2668332.2668366

[15] Alzubaidi, A., Kalita, J.: 'Authentication of smartphone users using behavioral biometrics', *IEEE Commun. Surv. Tutor.*, 2016, **18**, (3), pp. 1998–2026

[16] Rybnicek, M., Lang-Muhr, C., Haslinger, D.: 'A roadmap to continuous biometric authentication on mobile devices'. 2014 Int. Wireless Communications and Mobile Computing Conf. (IWCMC), Nicosia, Cyprus, 2014, pp. 122–127

[17] Alotaibi, S., Furnell, S., Clarke, N.: 'Transparent authentication systems for mobile device security: a review'. 2015 10th Int. Conf. for Internet Technology and Secured Transactions (ICITST), London, UK, 2015, pp. 406–413

[18] Meng, W., Wong, D.S., Furnell, S., *et al.*: 'Surveying the development of biometric user authentication on mobile phones', *IEEE Commun. Surv. Tutor.*, 2015, **17**, (3), pp. 1268–1293

[19] Gupta, S., Buriro, A., Crispo, B.: 'Demystifying authentication concepts in smartphones: ways and types to secure access', *Mob. Inf. Syst.*, 2018, **2018**, pp. 1–16

[20] Buriro, A., Akhtar, Z., Crispo, B., *et al.*: 'Mobile biometrics: towards a comprehensive evaluation methodology'. 2017 Int. Carnahan Conf. on Security Technology (ICCST), Madrid, Spain, 2017, pp. 1–6

[21] Li, S.Z., Jain, A.K.: '*Handbook of face recognition*' (Springer Publishing Company, Incorporated, USA, 2011, 2nd edn.)

[22] Technical Committee ISO/IEC JTC 1/SC 37 Biometrics, '*Information technology – vocabulary – part 37: biometrics*' (International Organization for Standardization, Geneva, CH, 2017)

[23] Hastie, T., Tibshirani, R., Friedman, J.: *'The elements of statistical learning', Springer Series in Statistics* (Springer New York Inc., New York, NY, USA, 2001)

[24] Anjos, A., Marcel, S.: 'Counter-measures to photo attacks in face recognition: a public database and a baseline'. 2011 Int. Joint Conf. on Biometrics (IJCB), Washington DC, USA, 2011, pp. 1–7

[25] Chingovska, I., Anjos, A., Marcel, S.: 'On the effectiveness of local binary patterns in face anti-spoofing'. 2012 BIOSIG - Proc. of the Int. Conf. of Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2012, pp. 1–7

[26] Li, S.Z., Jain, A. (Eds.): '*Open-set identification*' (Springer US, Boston, MA, 2009), pp. 1022–1022. Available at https://doi.org/10.1007/978-0-387-73003-5_805

[27] Technical Committee ISO/IEC JTC 1/SC 37 Biometrics, '*Information technology – biometric performance testing and reporting – part 1: principles and framework*' (International Organization for Standardization, Geneva, CH, 2006)

[28] Zhang, D., Xu, Y., Zuo, W.: '*Discriminative learning in biometrics*' (Springer, Singapore, 2016). Available at https://books.google.co.uk/books?id=EJRjDQAAQBAJ

[29] Dunstone, T., Yager, N.: '*Biometric system and data analysis: design, evaluation, and data mining*' (Springer, US, 2008). Available at https://books.google.co.uk/books?id=HXtagjFVEyIC

[30] Witten, I.H., Frank, E., Hall, M.A.: '*Data mining: practical machine learning tools and techniques*' (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2011, 3rd edn.)

[31] Popovici, V., Thiran, J., Bailly-Bailliere, E., *et al.*: 'The BANCA database and evaluation protocol'. 4th Int. Conf. on Audio-and Video-Based Biometric Person Authentication, Guildford, UK, 2003, vol. 2688, pp. 625–638

[32] Lui, Y.M., Bolme, D., Phillips, P.J., *et al.*: 'Preliminary studies on the good, the bad, and the ugly face recognition challenge problem'. 2012 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 2012, pp. 9–16

[33] European Union: 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general data protection regulation)', *Off. J. Eur. Union*, 2016, **L119**, pp. 1–88. Available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[34] Saeed, K.: '*New directions in behavioral biometrics*' (CRC Press, Inc., Boca Raton, FL, USA, 2016, 1st edn.)

[35] Technical Committee ISO/IEC JTC 1/SC 37 Biometrics, '*Information technology – biometric performance testing and reporting – part 2: testing methodologies for technology and scenario evaluation*' (International Organization for Standardization, Geneva, CH, 2007)

[36] Jain, A.K., Nandakumar, K., Ross, A.: '50 years of biometric research: accomplishments, challenges, and opportunities', *Pattern Recognit. Lett.*, 2016, **79, (Suppl. C)**, pp. 80–105. Available at http://www.sciencedirect.com/science/article/pii/S0167865515004365

[37] Martinez, C., Hollister, J., Marwick, B., *et al.*: 'Reproducibility in science'. Available at http://ropensci.github.io/reproducibility-guide/, accessed 22 January 2018

[38] Eagle, N., (Sandy) Pentland, A.: 'Reality mining: sensing complex social systems', *Pers. Ubiquit. Comput.*, 2006, **10**, (4), pp. 255–268. Available at http://dx.doi.org/10.1007/s00779-005-0046-3

[39] Torres, J., de los Santos, S., Alepis, E., *et al.*: 'Behavioral biometric authentication in android unlock patterns through machine learning'. Proc. of the 5th Int. Conf. on Information Systems Security and Privacy, ICISSP 2019, Prague, Czech Republic, 23–25 February 2019, pp. 146–154. Available at https://doi.org/10.5220/0007394201460154

[40] Shi, E., Niu, Y., Jakobsson, M., *et al.*: 'Implicit authentication through learning user behavior'. Proc. of the 13th Int. Conf. on Information Security. ISC'10, Boca Raton, FL, USA, 2010, pp. 99–113. Available at http://dl.acm.org/citation.cfm?id=1949317.1949329

[41] De Luca, A., Hang, A., Brudy, F., *et al.*: 'Touch me once and i know it's you!: implicit authentication based on touch screen patterns'. Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, Austin, TX, USA, 2012, pp. 987–996

[42] Feng, T., Liu, Z., Kwon, K.A., *et al.*: 'Continuous mobile authentication using touchscreen gestures'. 2012 IEEE Conf. on Technologies for Homeland Security (HST), Greater Boston, MA, USA, 2012, pp. 451–456

[43] Kolly, S.M., Wattenhofer, R., Welten, S.: 'A personal touch: recognizing users based on touch screen behavior'. Proc. of the Third Int. Workshop on Sensing Applications on Mobile Phones, Toronto, Canada, 2012, p. 1

[44] Sae-Bae, N., Ahmed, K., Isbister, K., *et al.*: 'Biometric-rich gestures: a novel approach to authentication on multi-touch devices'. Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, Austin, TX, USA, 2012, pp. 977–986

[45] Zheng, N., Bai, K., Huang, H., *et al.*: 'You are how you touch: user verification on smartphones via tapping behaviors'. 2014 IEEE 22nd Int. Conf. on Network Protocols, Raleigh, NC, USA, 2014, pp. 221–232

[46] Damopoulos, D., Kambourakis, G., Gritzalis, S.: 'From keyloggers to touchloggers: take the rough with the smooth', *Comput. Secur.*, 2013, **32**, pp. 102–114. Available at http://www.sciencedirect.com/science/article/pii/S0167404812001654

[47] Li, L., Zhao, X., Xue, G.: 'Unobservable re-authentication for smartphones'. NDSS, 2013, pp. 1–16

[48] Zhao, X., Feng, T., Shi, W.: 'Continuous mobile authentication using a novel graphic touch gesture feature'. 2013 IEEE Sixth Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 2013, pp. 1–6

[49] Antal, M., Szabó, L.Z., Bokor, Z.: 'Identity information revealed from mobile touch gestures', *Stud. Univ. Babes-Bolyai Inf.*, 2014, **59**, pp. 5–14

[50] Antal, M., Bokor, Z., Szabó, L.Z.: 'Information revealed from scrolling interactions on mobile devices', *Pattern Recognit. Lett.*, 2015, **56**, pp. 7–13

[51] Feng, T., Yang, J., Yan, Z., *et al.*: 'TIPS: context-aware implicit user identification using touch screen in uncontrolled environments'. Proc. of the 15th Workshop on Mobile Computing Systems and Applications. Hot-Mobile '14, Santa Barbara, CA, USA, 2014, pp. 9:1–9:6. Available at http://doi.acm.org/10.1145/2565585.2565592

[52] Sitová, Z., Šeděnka, J., Yang, Q., *et al.*: 'HMOG: new behavioral biometric features for continuous authentication of smartphone users', *IEEE Trans. Inf. Forensics Sec.*, 2016, **11**, (5), pp. 877–892

[53] Khare, K., Moh, T.S.: 'Mobile gesture-based iPhone user authentication'. 2015 IEEE Int. Conf. on Big Data (Big Data), Santa Clara, CA, USA, 2015, pp. 1615–1621

[54] Lin, C.C., Chang, C.C., Liang, D.: 'An approach for authenticating smartphone users based on histogram features'. 2015 IEEE Int. Conf. on Software Quality, Reliability and Security, Washington, DC, USA, 2015, pp. 125–130

[55] Neal, T.J., Woodard, D.L., Striegel, A.D.: 'Mobile device application, bluetooth, and wi-fi usage data as behavioral biometric traits'. 2015 IEEE 7th Int. Conf. on Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, USA, 2015, pp. 1–6

[56] Roy, A., Halevi, T., Memon, N.: 'An hmm-based multi-sensor approach for continuous mobile authentication'. MILCOM 2015 - 2015 IEEE Military Communications Conf., Tampa, FL, USA, 2015, pp. 1311–1316

[57] Wu, J.S., Lin, W.C., Lin, C.T., *et al.*: 'Smartphone continuous authentication based on keystroke and gesture profiling'. 2015 Int. Carnahan Conf. on Security Technology (ICCST), Taipei, Taiwan, 2015, pp. 191–197

[58] Yang, J., Li, Y., Xie, M.: 'Motionauth:motion-based authentication for wrist worn smart devices'. 2015 IEEE Int. Conf. on Pervasive Computing and Communication Workshops (PerComWorkshops), St. Louis, MO, USA, 2015, pp. 550–555

[59] Abate, A.F., Nappi, M., Ricciardi, S.: 'Smartphone enabled person authentication based on ear biometrics and arm gesture'. 2016 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 2016, pp. 003719–003724

[60] Abate, A.F., Nappi, M., Ricciardi, S.: 'I-Am: implicitly authenticate me person authentication on mobile devices through ear shape and arm gesture', *IEEE Trans. Syst. Man Cybern., Syst.*, 2017, **PP**, (99), pp. 1–13

[61] Al-Naffakh, N., Clarke, N., Dowland, P., *et al.*: 'Activity recognition using wearable computing'. 2016 11th Int. Conf. for Internet Technology and Secured Transactions (ICITST), Barcelona, Spain, 2016, pp. 189–195

[62] Antal, M., Nemes, L.: 'The MOBIKEY keystroke dynamics password database: benchmark results', 2016 IEEE 6th International Conference on Consumer Electronics, Berlin (ICCE-Berlin), 2016, pp. 35–46

[63] Inguanez, F., Ahmadi, S.: 'Securing smartphones via typing heat maps'. 2016 IEEE 6th Int. Conf. on Consumer Electronics – Berlin (ICCE-Berlin), Berlin, Germany, 2016, pp. 193–197

[64] Kumar, R., Phoha, V.V., Serwadda, A.: 'Continuous authentication of smartphone users by fusing typing, swiping, and phone movement patterns'. 2016 IEEE 8th Int. Conf. on Biometrics Theory, Applications and Systems (BTAS), Niagara Falls, NY, USA, 2016, pp. 1–8
[65] Phillips, M.E., Stepp, N.D., Cruz-Albrecht, J., *et al.*: 'Neuromorphic and early warning behavior-based authentication for mobile devices'. 2016 IEEE Symp. on Technologies for Homeland Security (HST), Boston, MA, USA, 2016, pp. 1–5
[66] Anjomshoa, F., Kantarci, B., Erol-Kantarci, M., *et al.*: 'Detection of spoofed identities on smartphones via sociability metrics'. 2017 IEEE Int. Conf. on Communications (ICC), Paris, France, 2017, pp. 1–6
[67] Anjomshoa, F., Aloqaily, M., Kantarci, B., *et al.*: 'Social behaviometrics for personalized devices in the internet of things era', *IEEE Access*, 2017, **5**, pp. 12199–12213
[68] Yao, F., Yerima, S.Y., Kang, B., *et al.*: 'Continuous implicit authentication for mobile devices based on adaptive neuro-fuzzy inference system'. 2017 Int. Conf. on Cyber Security And Protection Of Digital Services (Cyber Security), London, UK, 2017, pp. 1–7
[69] Fridman, L., Weber, S., Greenstadt, R., *et al.*: 'Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location', *IEEE Syst. J.*, 2017, **11**, (2), pp. 513–521
[70] Papavasileiou, I., Smith, S., Bi, J., *et al.*: 'Gait-based continuous authentication using multimodal learning'. 2017 IEEE/ACM Int. Conf. on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Philadelphia, PA, USA, 2017, pp. 290–291
[71] Shen, C., Li, Y., Chen, Y., *et al.*: 'Performance analysis of multi-motion sensor behavior for active smartphone authentication', *IEEE Trans. Inf. Forensics Sec.*, 2017, **PP**, (99), pp. 1–1
[72] Smith-Creasey, M., Rajarajan, M.: 'Adaptive threshold scheme for touchscreen gesture continuous authentication using sensor trust'. 2017 IEEE Trustcom/BigDataSE/ICESS, Sydney, Australia, 2017, pp. 554–561
[73] Attaullah, B., Crispo, B., Conti, M.: 'Answerauth: a bimodal behavioral biometric-based user authentication scheme for smartphones', *J. Inf. Secur. Appl.*, 2019, **44**, pp. 89–103
[74] Li, G., Bours, P.: 'A novel mobilephone application authentication approach based on accelerometer and gyroscope data'. 2018 Int. Conf. of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 2018, pp. 1–4
[75] Teh, P.S., Zhang, N., Teoh, A.B.J., *et al.*: 'TDAS: a touch dynamics based multi-factor authentication solution for mobile devices', *Int. J. Pervasive Comput. Communi.*, 2016, **12**, (1), pp. 127–153. Available at https://doi.org/10.1108/IJPCC-01-2016-0005
[76] Leyfer, K., Spivak, A.: 'Continuous user authentication by the classification method based on the dynamic touchscreen biometrics'. 2019 24th Conf. of Open Innovations Association (FRUCT), Moscow, Russia, 2019, pp. 228–234
[77] Meng, W., Li, W., Jiang, L.,: 'SocialAuth: designing touch behavioral smartphone user authentication based on social networking applications', in Dhillon, G., Karlsson, F., Hedström, K., *et al.* (Eds.): '*IFIP International Conference on ICT Systems Security and Privacy Protection (IFIP SEC 2019)*', (Springer International Publishing, Lisbon, Portugal, 2019), pp. 180–193
[78] Yang, Y., Guo, B., Wang, Z., *et al.*: 'Behavesense: continuous authentication for security-sensitive mobile apps using behavioral biometrics', *Ad Hoc Netw.*, 2019, **84**, pp. 9–18. Available at http://www.sciencedirect.com/science/article/pii/S1570870518306899
[79] Li, S.Z., Jain, A. (Eds.): '*Generalization*' (Springer US, Boston, MA, 2009), pp. 664–664. Available at https://doi.org/10.1007/978-0-387-73003-5_619
[80] Günther, M., Cruz, S., Rudd, E.M., *et al.*: 'Toward open-set face recognition'. 2017 IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 573–582
[81] Li, F., Wechsler, H.: 'Open set face recognition using transduction', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, (11), pp. 1686–1697
[82] Li, F., Wechsler, H.: 'Open set recognition using transduction'. Google Patents, US Patent 7,492,943, 2009
[83] Wechsler, H.: '*Reliable face recognition methods: system design, implementation and evaluation*' (Springer Publishing Company, Incorporated, USA, 2010, 1st edn.)
[84] Scheirer, W.J., de Rezende Rocha, A., Sapkota, A., *et al.*: 'Toward open set recognition', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013, **35**, (7), pp. 1757–1772
[85] Samangouei, P., Patel, V.M., Chellappa, R.: 'Attribute-based continuous user authentication on mobile devices'. 2015 IEEE 7th Int. Conf. on Biometrics Theory, Applications and Systems (BTAS), Arlington, VA, USA, 2015, pp. 1–8
[86] Shi, W., Yang, J., Jiang, Y., *et al.*: 'Senguard: passive user identification on smartphones using multiple sensors'. 2011 IEEE 7th Int. Conf. on Wireless and Mobile Computing, Networking and Communications (WiMob), Shanghai, People's Republic of China, 2011, pp. 141–148
[87] Android: 'Position sensors'. Available at https://developer.android.com/guide/topics/sensors/sensors_position.html, accessed 16 December 2017
[88] Android: 'Computing the device's orientation'. Available at https://developer.android.com/guide/topics/sensors/sensors_position.html#sensors-pos-orient, accessed 16 December 2017
[89] Android: 'Using the geomagnetic field sensor'. Available at https://developer.android.com/guide/topics/sensors/sensors_position.html#sensors-pos-mag, accessed 16 December 2017
[90] Android: 'Using the gyroscope'. Available at https://developer.android.com/guide/topics/sensors/sensors_motion.html#sensorsmotion-gyro, accessed 16 December 2017
[91] Android: 'Using the gyroscope'. Available at https://developer.android.com/guide/topics/sensors/sensors_motion.html, accessed 16 December 2017
[92] Innoventions, I.: 'Gravity sensor in smartphones and tablets'. Available at https://www.rotoview.com/magnetometer.htm, accessed 11 September 2019
[93] Gunes Kayacik, H., Just, M., Baillie, L., *et al.*: 'Datadriven authentication: on the effectiveness of user behaviour modelling with mobile device sensors', ArXiv e-prints, 2014
[94] Moya, M.M., Koch, M.W., Hostetler, L.D.: 'One-class classifier networks for target recognition applications', United States, 1993
[95] Ross, D.: 'Processing biometric data? be careful, under the GDPR', 31 October 2017. Available at https://iapp.org/news/a/processing-biometric-data-be-careful-under-the-gdpr/, accessed 4 May 2018
[96] Dunlop, A.: 'GDPR: personal data and sensitive personal data', 21 November, 2016. Available at https://www.burges-salmon.com/news-and-insight/legal-updates/gdpr-personal-data-and-sensitive-personal-data/, accessed 5 May 2018
[97] Irwin, L.: 'GDPR: things to consider when processing biometric data', 15 September 2017. Available at https://www.itgovernance.eu/blog/en/gdpr-things-to-considerwhen-processing-biometric-data, accessed 4 May 2018
[98] Stodden, V., Borwein, J., Bailey, D.H.: 'Setting the default to reproducible', *Somputational Sci. Res. SIAM News*, 2013, **46**, (5), pp. 4–6