POZNAŃ UNIVERSITY
OF ECONOMICS
AND BUSINESS

Piotr Kałużny

Behaviour-based user authentication for financial services

Behawioralne uwierzytelnianie użytkowników usług sektora finansowego

Doctoral dissertation

PhD Supervisor: prof. dr hab. Witold Abramowicz

Auxiliary PhD Supervisor: dr hab. Agata Jolanta Filipowska

Date of submission:

Supervisor's signature

Poznań 2021

Contents

Chapter 1

Introduction

1.1 Motivation

The value of the global financial market is a significant part of the world's economy. At the end of 2020 the global financial assets climbed to 404 trillion USD: 38,5% belonged to the commercial banks, 7,5% to the central banks, 4,5% the public financial institutions and 49,5% to the NBFI (non banking financial institutions) sector, which includes insurance corporations, pension funds, OFIs and financial auxiliaries (Financial Stability Board, 2020). The NBFI sector has grown faster than the banking sector over the past decade. The overall averaged growth of the financial sector for years 2013-2018 was estimated at 5.2%, and over 6.6% in year 2019. Despite the losses caused by the COVID-19 crisis, some parts of the sector still experience growth, especially in the domain of new technologies in the financial sector exemplified by FinTechs. This market has continued to help expand access to services during the COVID-19 pandemic — particularly in emerging markets — with strong growth in all types of digital financial services except lending, according to a joint study by the World Bank, the Cambridge Centre for Alternative Finance at the University of Cambridge's Judge Business School, and World Economic Forum (www.worldbank.org, 2020).

The value of the banking sector in Poland in September 2019 achieved over 2 billion PLN (KNF, 2020) according to the Polish Financial Supervision Authority (KNF). This value does not include a large portion of other financial services and new FinTech startups which work in the same sector. This value is significant in terms of the Polish and world's economy, and the underlying processes and changes are of the highest importance for the economics and finance research. The potential technological and organizational challenges of providing said services

are also demanding from the perspectives of actors directly involved, be it banks, financial institutions or other companies.

Unfortunately, despite the rapid development of technology allowing for better processing of information between the sector's participants, it is also subject to multitude of risks, which highly influence the stability of its core institutions. Significant portion of them is connected with the cash flow in this sector and burdened by the losses caused by frauds. The total number of non-cash payments in the euro area was estimated in 2019 by the European Central Bank (ECB) at 98 billion. With the total value of €162.1 trillion it is an increase of 8.1% from 2018. In Europe, frauds using cards issued within SEPA (Single Euro Payments Area) worldwide amounted to €1.80 billion in 2018 (European Central Bank, 2019) as presented in the Figure 1.1. The scale on the left of the figure presents the total value of the frauds (in EUR millions), while the right scale showcases the value of frauds as a share of value of transactions (percentages). Overall, the value of frauds from card-not-present (CNP - i.e. payments via the Internet, mail or phone) payments amounted to 79%. CNP fraud accounted for €1.43 billion in fraud losses in 2018 (an increase of 17.7% compared with 2017). In comparison to the point of sale (POS) and ATM frauds they accounted respectively only for 15% and 6% of the overall SEPA frauds value. The rising value of frauds where card is not physically present is an important issue for the financial sector. Its goal should be to minimize those fraud values to create better market conditions and guarantee stability. When considering the growth of such transactions from 2012, CNP frauds are especially dangerous, and ECB reports estimate a steep increase in frauds with the use of mobile devices. This leaves a large problem for today's economy aimed at the financial sector, which should minimize those fraud values to create better market conditions and guarantee its stability.

Current changes in banking and payment systems are influenced highly by the development of technology and ubiquity of mobile services. According to the recent GSMA report (GSMA, 2019) about 5 billion people all over the world use mobile phones (67% of the world population). Nearly 60% of those devices fall into smartphone category, and 66% of those people accesses the Internet with their mobiles. This ubiquity makes those devices a basic tool for a variety of everyday tasks, including mobile payment and banking services. Banks and financial institutions need to adapt to the needs of their customers, and those include mobile banking applications. Use of smartphones and mobiles for financial services is not a novelty but a requirement, as in 2019 nearly 90% of users used their phones for such services (Meola, 2019).

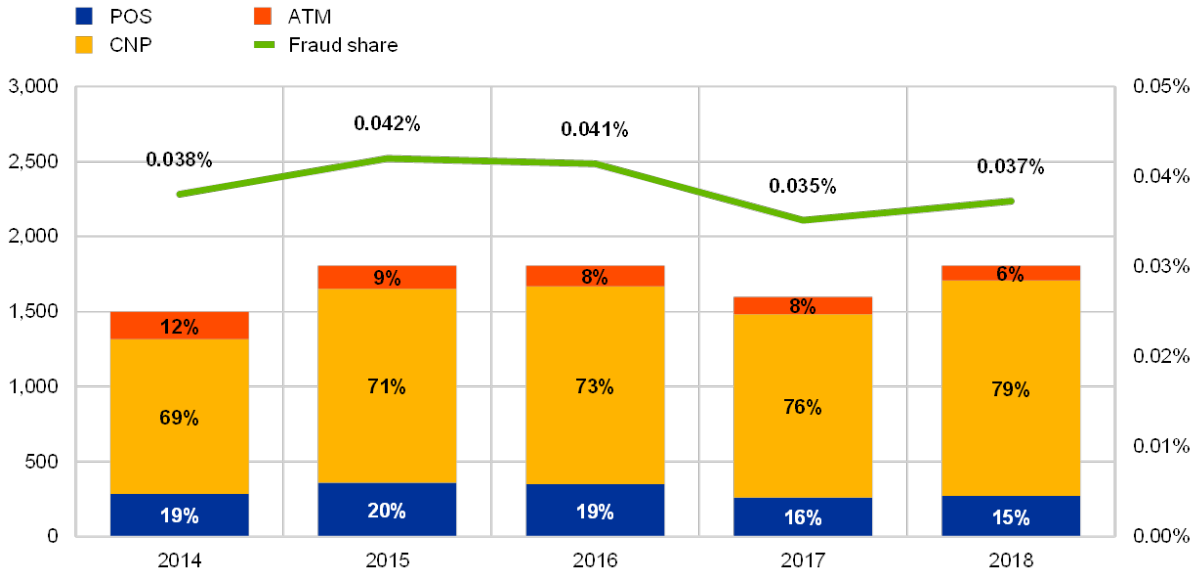


Figure 1.1. Evolution of the total value of card fraud using cards issued within SEPA from 2014-2018. Source: (European Central Bank, 2019)

The emergence of new payment systems and applications such as BLIK[1], PayU[2], Revolut[3], Apple Pay[4] or Google Pay[5] creates requirements for building banking systems which are interoperable with those solutions and their mobile-centric environment. Visa „Digital Payment 2017" European market study showcased that 77% overall and 86% among 18-34 years respondents consider themselves "Mobile Money users" (Visa, 2017b). Similarly, according to the Polish survey conducted a year before, among university students and workers, 88% of members of these groups utilized mobile banking services (Staszczyk, 2016).

Those results prove that more and more users use mobile devices for accessing the financial services. Due to that, the traditional model of physical branches in financial institutions gives way to the new mobile model. As customers not only check their account balance, but also conduct payments and buy new services, their demand for access to all of the sector's services is only increasing. The Deloitte 2018 banking outlook report enumerates the main challenges for the transformation of current banking models and strongly points out the model of "mobile-centric" and "customer-centric" banking as a model for the future (Srinivas et al., 2018). This model, presented in the Figure 1.2 is shaping new banking strategies, as they need to adapt to this new environment. This process however is a significant challenge for the sector, as tradi-

[1] https://blik.com/
[2] http://payu.com/
[3] https://www.revolut.com/
[4] https://www.apple.com/en/apple-pay/
[5] https://pay.google.com/intl/en_en/about/

tional forms of confirming the identity, such as signature, ID verification or even photo comparison of the customer with the government ID photography supplied is non applicable in those mobile scenarios. This ever-increasing customer demand for mobile services forces financial institutions to search for new ways of authentication for their users that may help or even entirely replace existing solutions such as passwords and card verification codes. Changes in the payment systems also adapt to this new paradigm of mobile-centric users and are influenced by the intense development of new technologies in this field. Appearance of various payment brokers on the market confirm the dynamic changes happening in the financial sector.

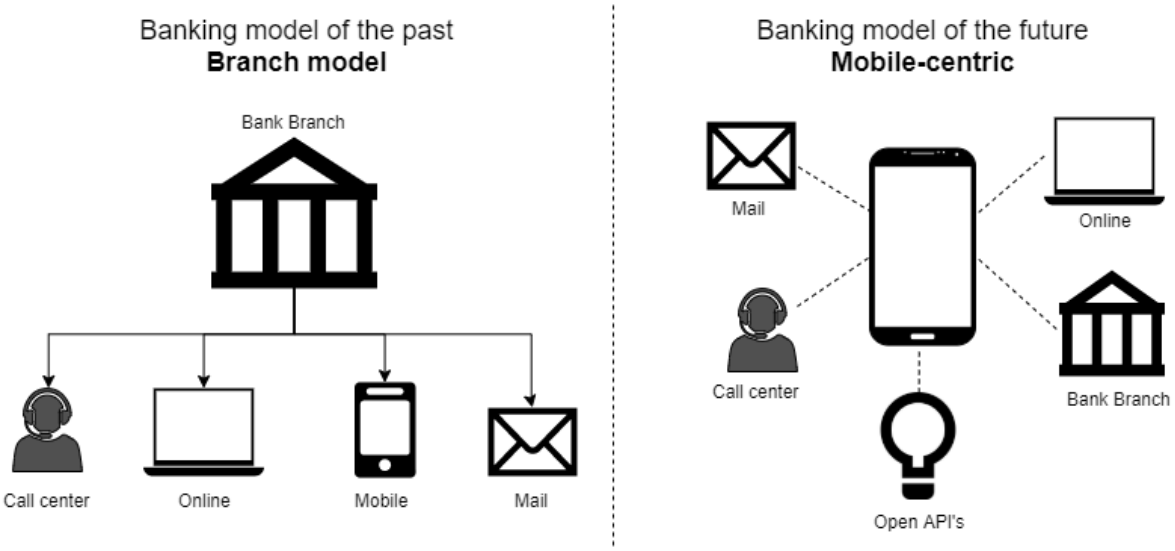


Figure 1.2. New mobile-centric banking model overlook. Source: own development based on (Srinivas et al., 2018).

The appearance of multiple service providers and intermediaries offers different types of added value to traditional services, from accelerating the payment process to managing finances of a person or institution. This however threatens the banks position in the sector. Even when the first contact with the customer is handled by an intermediary financial service, banks and credit card providers have a responsibility of ensuring the transaction security. As institutions of public trust, banks still retain more customer confidence in this matter than their competition (Świeszczak, 2017) and this is their main source of competitive advantage. Adoption of their services to better suit this new mobile-centric environment is mandatory, so they can retain it. The institutions which will adopt faster may collaborate with FinTechs in areas which would be hard to invest in for banks due to their limited elasticity and slow adoption of new standards. This strategy is inline with Deloitte findings, but over time may be a necessity more than an opportunity. New EU Payment Services Directive (known widely

as PSD2)(Europen Comission, 2015) requires banks to allow for accessing their services by external providers as an ”Open Banking” standard. The institutions which will adjust reluctantly may loose opportunities to lead and quickly adopt to new mobile financial services trends and users' needs.

Above mentioned studies present that customers of banks and financial institutions are not able to assess how safe and secure are the services used by them. The issue of security awareness in the field of mobile applications has been researched since the appearance of mobile phones (Imgraben et al., 2014; Mazhelis, 2007). As there is an increasingly large portion of customers of mobile services, their trust in mobile solutions increases (Visa, 2017b), while their security awareness may not (also due to the complexity of used technologies). VISA (Visa, 2017b) and Mastercard (Mastercard, 2017) studies provide some insights as to why users are using mobile solutions for accessing financial services. The users point out convenience as one of the main reasons for using mobile banking and do not want to carry cash, but are also interested in security of offered services. From those customers, most clients utilize mobile devices to keep track of their finances and pay for everyday services using mobile banking and e-wallet applications. This dynamics is even more prevalent in Poland, where BLIK[6] payment service allows customers to pay by providing short one-time codes and transfer money utilizing phone numbers. The Polish customers also use mobile payments regardless of the transaction value (Visa, 2016). While convenience is the main reason why users choose mobile wallets, 37% of them believes that safety and security are also an important benefit of e-wallet services.

Following the requirements of the customers, who demand usable and secure services, the financial sector needs to adapt to the technological environment of the mobile applications. The above mentioned user needs, high competitiveness level on the financial services market (with the new FinTech and mobile payment options appearing) and banking main goals for providing safe and secure services lead to the conclusion that banks and financial institutions need new solutions which can provide services better suited for this environment. With users conducting everyday payments by mobile banking applications, the new mobile-centric banking model becomes a goal to which all modern financial institutions try to adapt (Deloitte Center for Financial Services, 2018). The solutions provided in this mobile-centric model need to include the newest technologies to adapt to their customers and provide a compromise of security and convenience, which can help them to meet their users' needs. In this new en-

[6] https://blikmobile.pl/en/

vironment financial institutions must provide services which are not only convenient, but also secure (Visa, 2016) to meet their customer requirements, stabilize their position on the market and integrate with new FinTech solutions. Those traits can be ensured by the use of biometric authentication services which are already available in some of the banking applications e.g., utilizing fingerprint scanners. The use of biometrics factors and methods allows for providing high convenience and security of provided services (Zakonnik & Czerwonka, 2014). On the other hand, the security of the biometric pattern (such as fingerprint scan), shared by the customer with more and more service providers, remains an issue and could be dangerous in cases when it would leak out in some security breach.

While ensuring the security of a provider's application, what is also important is the security of the device itself. Not all mobile devices are equipped with the required biometric sensors of sufficient quality and about 40% of users do not secure their device in any way - be it PIN or biometrics (Fridman et al., 2017). This poses a significant threat in securing the transactions and preventing frauds. Service providers need not only to make their services available in the mobile channel, but also take into consideration that the device might have been compromised.

This might happen either due to theft or acquisition of a user's phone, but also from malware installed on the device. According to the broad study of malicious applications targeting financial applications from 2018 (Black et al., 2018), the rough categorization of behaviour included: persistence (working in the background), configuration (firing when user visits a specific website), process injection (undetectability from benign processes), information stealing (of for example user passwords), injection (of attacker data into the banking form), network communications (intercepting user credentials and communication with the service), backconnect (allowing root access), screenshot and video capture, and anti-analysis. With such a broad catalogue of potential threats it is a problem to make sure tha the user is also aware of the action and initiated the process. While transaction may be performed from the user's device or with the use of proper password, it could still be fraudulent and unauthorized, being done by a malware or by an unauthorized attacked. Confirming that the user consciously carried out the action is the issue solved by proof of presence authentication (Samet et al., 2019)[7]. Currently, banks utilize mostly static and transactional data for their anti-fraud systems. With mobile devices they can enrich these systems with behavioural data generated by the devices, not only directly during the transaction, but also through the whole process of interaction with a mo-

[7]See glossary in Appendix A.

bile application. Inline with the assumptions of behavioural economics, it is crucial to analyse real behaviour of business entities and customers in terms of their psychological motivation and captured actions. The principles of the behavioural economics do not assume anything, but rather study the captured behaviour of those entities (Polowczyk, 2009). Results of such observations are already included in current anti-fraud systems, describing e.g., what types of transactions the user performed, where, and when. Broadening the potential of data-driven analysis in this area may offer improvements to combat the aforementioned fraud issue in the mobile channel. From all of the sub areas of behavioural economics this works broadens the concept of experimental economics and behavioural finance. The first area is the most visible in various research experiments relying on the data used to validate the proposed method. On the other hand, this work is a part of behavioural finance mostly because it studies the nature of financial decisions of customers with regards to their account security in mobile financial applications.

Current authentication of transactions (and also every possible action) in mobile financial applications can be simplified to a three factor model, consisting of: knowledge, possession and inherence factors (Renaud, 2005). Passwords and PIN numbers (representing knowledge factor) are used most commonly to authenticate a user. Yet they work as an ”all-or-nothing” (Hayashi et al., 2012) authorization, where either complex pattern needs to be used at all time, which is hard to remember for users, or we risk using a simple pattern, which can be easily guessed by a perpetrator. Some banks apply different passwords for different services (e.g., Millenium Bank in Poland), in which one is used for transaction of higher value and connected with more risk. This model is becoming a problem for users, which tend to use very simple and predictable passwords or reuse them. This becomes an issue as clients need to remember multiple passwords for multiple services (Bonneau & Preibusch, 2010). The use of a token factor can allow for simpler passwords, as it is used in one time passwords (OTPs), generated in services like BLIK or provided via SMS. But the generation of such token still relies on a security of the user device itself, which may be compromised by malware mentioned above. There also exist an issue of an insider threat (see glossary A) (Hayashi et al., 2012; Muslukhov et al., 2013), when a person close to the original possessor of a device, having access to it and sometimes even credentials, may intentionally or unintentionally perform a transaction. This insider threat is an authentication concept (and not the more widely known definition used in economics), and may include user family, spouse or children performing a transaction without the account

holder's knowledge. This may be especially prevalent when children perform unauthorized card transactions or payments using their parent's mobile phone, through mobile games and similar applications.

The inherence factor - mainly exemplified by fingerprint biometrics is an easy to use and secure method for authentication, but relies on a sensor installed on a device. And while in many developing countries, not all devices may have a sufficient sensor, recent Apple iPhones 8 are not equipped with them, relying on a face detection technology instead. The use of face biometrics also remains an alternative to fingerprint biometrics, but for now it is available with satisfying accuracy only on a very small number of devices - namely, as of 2020 produced only by Samsung and Apple. Due to that, developing biometric solutions which are not tied to a specific company hardware or software, and may be available on multiple devices remains an open issue. For physical biometrics, the theft of a pattern and spoofing (by a fingerprint scan, mold or face photo) are a significant problem. The permanence of such patterns, as each human has one face and 10 hand fingerprints also makes them vulnerable for potential disclosure. As it is stolen, the pattern itself is compromised. This makes a secure storing and transfer of the pattern an issue to be solved in financial applications.

As a summary of the above mentioned analysis, a list of current problems the financial sector is facing can be identified:

1. **Banks' position as institutions of public trust** - customers trust the banks and are confident when bestowing them with their money or personal data (Staszczyk, 2016; Świeszczak, 2017). Other companies, such as payment brokers need to not break this trust to the sector itself, as it can cause a large loss of customers for all of the parties involved.
2. **Fraud detection** - financial sector needs methods that can effectively detect fraud, especially CNP fraud connected with mobile payments and banking applications (European Central Bank, 2019).
3. **Adoption of the mobile-centric model** - financial institutions need to develop services and transform their standard branch model of operation to mobile devices (Meola, 2019; Srinivas et al., 2018).
4. **Providing more advanced services in the mobile channel by banks and OFIs** - providing not only basic payment options, but also services which are burdened with a higher risk (Meola, 2019; Srinivas et al., 2018; Visa, 2017b).

5. **Increasing the usability of mobile applications** - enabling new functions or increasing security must be paired with constant improvement of usability to meet customers' needs (Mastercard, 2018; Visa, 2017b).
6. **Proving more flexible authentication model** - instead of all-or-nothing access the applications should allow access to low risk functions (like checking an account balance or low amount transfers (Samojło, 2019)) with less usability-hindering authentication procedures than those connected with high risk (Chatterjee et al., n.d.; Hayashi et al., 2012; Kałużny & Stolarski, 2019).
7. **Increasing the security of mobile financial applications** - to prevent potential cases of fraud, the overall security mechanisms used in the mobile financial applications (Mastercard, 2017; Vignolo, 2019) and the mobile phones must be improved (Fridman et al., 2015; McDonnell et al., 2014).
8. **Enhancing the risk assessment on transaction level** - financial institutions need additional information for improving the risk assessment of transactions (Vignolo, 2019), especially the ones conducted in the mobile channel as complying with the PSD 2 requirements (Europen Comission, 2015).
9. **Securing the applications from malware** - either detection of malware on user's phone or rejection of malware-induced fraud attempts is important for mobile applications (Black et al., 2018; Kałużny, 2019b; Lovisotto et al., 2017).
10. **Protecting users from somebody accessing the mobile application through stolen credentials** - whether the attacking party is unknown to a user and the credentials were accessed by phishing, or the access is done by somebody that the user knows, either intentionally or as a result of a mistake (spouse, child) (Hayashi et al., 2012; Milton & Memon, 2016; Muslukhov et al., 2013).
11. **Compliance with the PSD 2 requirements** - enabling two-factor authentication in a manner acceptable by both the legal requirements of PSD 2 (Europen Comission, 2015) and users. Allowing that while providing high usability and maintaining required security level is a challenge for financial institutions, as most of the time more secure methods lower usability and may cause user dissatisfaction. On the other hand, financial institutions must make sure to minimize the transaction fraud risks.
12. **Development of a platform for the assessment and confirmation of user identity for Open Banking** - as more and more FinTech companies are utilizing bank architectures,

customers may not want to entrust them with their personal and biometric data. The trusted and larger participants of the market, which are banks may be required to hold this information and confirm user identity even on transaction level for smaller participants such as FinTechs. This may effectively make banks the holders of customer digital identity in the sector, which requires good identification and authentication processes (Subasinghe, 2019).

13. **Protecting users privacy and their confidential data** - protection of customers personal data and other information which may fall under the General Data Protection Regulation (GDPR) is of high importance for all institutions in the sector. Compliance with such regulations is required not only due to the potential penalties, but also because the leak of any such information may influence not only the organization which allowed this to happen, but may compromise user trust to a specific technology or even trust to the entire financial sector. As proven by recent studies, spoofing attempts on traditional biometrics and even early keystroke behavioural biometrics are an increasingly relevant issue (Awad, 2017; Ramachandra et al., 2019; Wu et al., 2020).

Those issues are also connected with more technical problems, that are mentioned in the security domain (S) scientific literature and tied to mobile applications environment, including:

S.1 **Non point-of-entry or continuous mobile authentication** - allowing authorization with different levels of privileges, which is impossible in password/pin or traditional biometric authentication. The issue was discussed by multiple researchers and addressed by many algorithms (Fridman et al., 2017; F. Li et al., 2014; Patel et al., 2016; Wójtowicz & Joachimiak, 2016).

S.2 **Usability of the authentication processes** - improving the perceived usability, either by simplifying the authentication process or decreasing the number of times user password is required (Abuhamad et al., 2020; Kainda et al., 2010; Patel et al., 2016).

S.3 **Device security** - increase the security of the device, which requires no additional actions from a user and takes into consideration potential user negligence (Alotaibi et al., 2015; Lawless Research, 2016).

S.4 **Pattern security** - solves issues of potential pattern leak connected with permanence of biometric patterns (Bolle et al., 2013; Lovisotto et al., 2017).

S.5 **Malware detection** - detect malware which may impersonate a user and perform actions from his/her mobile phone (Lovisotto et al., 2017; Milton & Memon, 2016).

S.6 **Acceptance of new technologies by users** - use of complicated and new technologies is always connected with certain reluctance from users. Measuring the adoption and acceptance of used technology is important in terms of its wide deployment in services (Luo et al., 2010; Mills & Zheng, 2019; Sanjith, 2017).

S.7 **Device and vendor independent authentication** - not every phone is equipped with high quality biometric sensors. And as such, good authentication method should not rely on state of the art sensors and the user having access to the newest generation of mobile phone to utilize it.

S.8 **Non binary authorization levels** - standard password/physical biometric authentication mechanisms work as „point-of-entry" mechanism. Effectively, user is either fully authorized or not, this can be circumvented by using different patterns (PIN for logging, password for other actions) but is inefficient. New methods should allow for non binary authorization (Crawford & Renaud, 2014; Primo et al., 2014).

S.9 **Proof of presence authentication** - which could protect the users from unintentional, falsely authorized transactions and the authentication insider threat issue (Muslukhov et al., 2013).

S.10 **Transparency of the method** - with methods tied to the specific hardware producers like Apple and Samsung measuring the unbiased level of method's performance and error rate becomes difficult. This in turn puts even more emphasis on scientific community on evaluation of the authentication methods in various scenarios to ensure their constant improvement.

Both of those lists summarise problems which are highly connected with the challenge of satisfying the security versus usability in the mobile channel along with providing mobile financial services that can protect users from private information leaks and fraud. This is also presented in the Table 1.1, which classifies issues described into separate topics. Users tend not to prefer the most secure systems which are of low usability, and often find alternative ways of interaction or avoid the system completely (Kainda et al., 2010). Most of the users choose the most usable method which will be allowed from a security perspective, providing an acceptable level of security, not the highest. This issue highly influences the possibility of ensuring safety of transactions and preventing frauds in the financial sector. The method using very long, unique password, additionally supplied with standard biometrics will not be used although it is available as of today. This happens due to the usability decrease it is connected

with. These issues create a need for developing methods which increase the security without hindering the usability of the authentication process.

Table 1.1. Description of the financial sector issues.

Category	Financial sector issue	Security domain and authentication issue
Security	7, 8, 9, 10, 11	S.3, S.4, S.5
Usability	5, 6	S.1, S.2 S.8
Privacy protection	10, 13	S.10
Fraud detection	2, 8	S.4, S.5, S.9
Mobile services adoption	3, 4, 5, 7, 9, 11	S.7
Banks role and technological challenges	1, 12	S.6, S.7

Source: own elaboration

1.2 Objective of the research and the thesis

The problems connected with the financial sector adoption of the mobile-centric model are diverse, but are indisputably tied together with the adoption of authentication and fraud detection methods in mobile-centric environment. Financial institutions face increasingly more fraud cases connected with card not present (CNP) transactions. Large part of them is done through mobile channel, due to the unauthorized access, phone theft, malware, phishing and resulting credentials theft. Assessing the identity of a user becomes an important issue in the mobile-centric environment, where no bank employee can verify the customer is who he or she claims to be. Legal requirements connected with PSD 2 and GDPR impose a multitude of obligations for not only securing the transactions but also ensuring data security, which is especially problematic in Open Banking environment where a limited level of trust to third parties such as FinTechs should be employed. Therefore, financial institutions must develop methods, suited for the mobile environment that would allow them to provide more sophisticated and diverse services through their applications while ensuring transaction and authentication security. Developing a reliable method, more suited to solve the aforementioned issues, providing not only safe, but also usable authentication processes is of the utmost importance for the sector. Shielding users from frauds, malware and use of stolen

credentials while retaining usability and protection of their confidential data are the new challenges which the method should address. Providing a method that can utilize behavioural information unique to the mobile device to enrich fraud detection systems on transaction level could help in minimizing losses caused by CNP frauds. These issues allowed the formulation of the following main research problem:

Financial services require authentication methods suited for mobile application environment, which could enrich current fraud-detection systems on a transaction level, while retaining the security and improving the usability of the process itself compared to the currently used methods.

Addressing this research problem, having in mind challenges and requirements of the financial sector presented in the previous section, we define the following research questions:

RQ1: *What are the requirements for the authentication method that can be used in mobile banking and payment applications from the perspectives of customers, providers (banks and financial institutions) and third parties?*

RQ2: *What sensors, methods and their combinations can be utilized by a mobile financial application authentication based on behavioural aspects?*

RQ3: *What features can be used to create behavioural patterns of a user?*

RQ4: *Can complementary features be chosen based on the defined characteristics and combined to design an authentication method which satisfies requirements of banking and payment scenarios?*

RQ5: *What scenarios can be used for benchmarking of a designed method and how to evaluate and validate the designed method's results in the mobile financial application?*

The goal of this research is to characterize the requirements for the new methods and services that can be utilized in mobile banking and payment applications and to create method that can improve the security and usability of current mobile banking and payment application authentication process. The importance and the business impact of this issue on the financial sector was stated and explained in the motivation section. The assumption this work relies upon, is that behavioural biometrics methods can be used as such improvement for increasing the usability of currently provided authentication process, partially by adjusting the authentication processes to the mobile environment. The use of those methods however pertains to multiple questions about their performance, influence on usability and scenarios of use in mo-

bile applications. While those methods may be promising, they have not widely been adopted as of today. Secondly, despite the multiple examples of implementations in the literature they showcase highly varying results. These observed differences in some cases contradict the notion that these methods could be safely used for ensuring transaction and account security in mobile financial services environment. However, finding a right method, supplied by data from some of multiple sensors installed on today's smartphone may accelerate the transformation to mobile-centric services and lower the fraud value, while also enabling authentication for new financial services available in the channel. To design such a method an analysis of literature on behavioural biometrics and most promising methods needs to be compared to sector's requirements and currently used authentication methods. Further on the method needs to be tested and evaluated in scenarios which will prove its use in financial sector applications. In the presented work a method for behaviour-based user authentication for financial services is designed, utilizing data collected from mobile device sensors.

The work focuses on characterizing the possible features which may be used in behavioural authentication in behavioural biometrics systems and evaluates their application in various authentication approaches. It also aims to discuss the evaluation methodologies and use cases in which behavioural biometrics may be utilized, along with possible benefits that may be achieved.

The thesis in this research is as follows:
An authentication method designed using behavioural biometrics can be deployed in a mobile financial application and achieve error rate lower than currently employed mobile face detection methods, providing higher usability.

Proving this thesis is guided by achieving the following research goals, being inline with previously defined research questions:

RG1 Building a model of requirements for the behavioural biometrics authentication method in financial environment.

RG2 Analysis of research results in the field of behavioural biometrics with the goal of choosing a method that fits the mobile financial application environment.

RG3 Characterizing features which may identify and allow to authenticate users based on the chosen method.

RG4 Design an artifact - an authentication method which can meet the accuracy and error rates criteria, while also satisfying other requirements for banking and payment scenario use.

RG5 Evaluate the designed method performance based on previously defined criteria and validate it in scenarios applicable for mobile financial applications to prove its feasibility in the chosen environment.

Table 1.2. Description of the research questions and goals correspondence.

Research question	Research goal	Corresponding chapter
RQ 1	RG 1	Requirements model presented in Chapter 2 Section 2.3.
RQ 2	RG 2	State of the art analysis of biometrics methods presented in Chapter 3, Section 3.4.
RQ 3	RG 3	Groups of features for each modality presented in Chapter 3, list of proposed features for the designed method is presented in Chapter 4 Section 4.4. Their importance across multiple datasets is analyzed in Chapter 5 Section 5.2.3.
RQ 4	RG 4	Method designed and described in Chapter 4.
RQ 5	RG 5	Results of the verification and validation presented in Chapter 5.

Source: own elaboration

Focusing on the above mentioned research and business problems, the goal of this research is to provide a contribution which enables a more flexible security of financial applications, more suited for the mobile environment. The problems handled by this work correspond to the following issues listed in the Table 1.1 which were identified for the sector:

- 3, 5, 7, 11, 12, S.1, S.2 - by providing a method that allows behavioural authentication on mobile devices without additional cost in terms of usability loss caused by the authentication process. The proposed design of the method will be compliant with the requirements of PSD2 representing a factor applicable in strong customer authentication (SCA)[8], while also being built for the mobile application in mind. By decreasing the number of times a user is asked for password/PIN or a fingerprint, the method will also provide a clear usability benefit in low-risk functions of those applications, while also working as an additional authentication factor for high value transactions.

[8]See glossary in the Appendix A.

- 2, 6, 8, S.8 - the method designed aims at providing probability estimates which allow for the assessment of behavioural pattern match of samples with the profile built previously. This in turn offers not only a more flexible approach to authorization and possible restriction of high risk functions based on the result of this comparison, but also enables the transfer of those estimates to the fraud-detection system for the assessment of risk. Risk management is then possible both in the authorization process on the device and in the provider's anti-fraud infrastructure.
- S.4 - due to the characteristics of a mobile phone sensors, behavioural biometrics can be to some extent unique to the device it has been captured on. This may be especially relevant in touchscreen biometrics, due to the different touch sensors and screen size, resolution and proportions. This uniqueness may be a hindrance for the identification process of a user that has changed his/her device, but lack of the pattern permanence decreases the risk of pattern theft.
- S.6 - the acceptance of a given technology must be studied with regard to the user requirements and main concerns. This dissertation aims at describing potential opportunities and drawbacks connected with the implementation of new methods. However, assessing user acceptance of a given instantiation of the model is out of the scope of this work.
- S.7 - relying on a set of sensors available on most or all of the devices, the method aims at solving the issue of fingerprint/face biometrics being viable and accepted by financial institutions only on a very limited set of actions, meaning only limited authorization can be achieved. Applying behavioural biometrics aims at allowing varying levels of privileges to the users based on the method's result.
- 10, S.9 - recognition of unauthorized access of an "insider" (friend, family or child having access to the device) (Muslukhov et al., 2013) will be studied in terms of gender and age group recognition accuracy provided by the method. This could incur additional security mechanism in those cases, while not hindering the owner user experience with the application.
- S.10 - by disclosing the method's design and results tested on multiple datasets, the method can be carefully analysed by security specialists. Providing new evaluation and validation scenarios the contribution to the security field of research may encourage new

researchers to further on improve the method and improve the overall security of authentication mechanism in mobile services.

Some issues may also be solved partially, while the method designed may offer some solution to the issue, it might not be evaluated in the dissertation or proven explicitly. For example, the method utilizes patterns generated by a user, which protect it from a malware trying to simulate user behaviour in the application (problem 9 and S.5). While malware attacks will not be studied empirically on a simulated scenario in this work, the design of the method itself may increase the difficulty level of performing such malware attacks by providing proof of presence authentication. Studying some of the issues that are not handled directly by this work, due to the limited size of the dissertation, may be a good starting point for directing the further work in the topic. The methodology behind the research follows the principles of the Design Science and the special guidelines for Information Systems (Hevner et al., 2004) apply as well. The artifacts developed during the process, inline with the dissertation topic, could cover:

- new or adopted and improved methods for behavioural biometrics authentication,
- development and evaluation of processes for potential enrichment of authentication with behavioural biometrics.

To answer these research questions a broad analysis of literature is required which describes the characteristics of behavioural biometrics. Its main goal is to provide preliminary evaluation of the behavioural biometrics methods to tackle the research problem defined and lead to the design of an authentication method - main artifact of the dissertation.

1.3 Research methodology

The paradigm of the Design Science Research (DSR) was selected as a leading methodology for research described in this dissertation. As the topic itself is placed at the intersection of Quantitative Economics, Finance and Computer Science, the special guidelines for Information Systems (Hevner et al., 2004) apply as well. This research paradigm is highly focused on problem solving. Design science in Information Systems (IS) pertains to the creation of artifacts to solve real life problems (Prat et al., 2014).

The research process begins in an environment. By searching for problems and opportunities (which may be understood as a research gap), one should take into account the application

domain, including people, organizational and technical systems already in place. In this case it is financial services environment. Those identified problems are not abstract and purely theoretical, as the environment gives them a clear context and leads to the identification of issues that should be addressed and are important and non trivial for the domain. The requirements and significance of the underlying issues of the research must be proved in the ”Relevance Cycle”, as the requirements appear. The proposed solution is not only inspired by its environment but given shape and meeting the requirements created by the domain. Additionally, this process leads to the design of scenarios for the validation of achieved results (artifacts) later on in the research process. Then the design process begins, by defining the constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implementations and prototypes of systems), where results of each are provided as artifacts. Those include new models, methods, frameworks or processes, which Design Science evaluates to solve identified problems, based on previously set requirements. The effects, according to the ”Rigor Cycle” must be grounded and proven by clear evaluation and validation, they must also contribute to the existing body of knowledge (KB). This in turn leads to the creation of highly applicable results (in form of an artifact), but still methodologically sound and innovative (in terms of clear addition to the KB) results.

Design Science is a paradigm that guides the approach and gives the researcher much freedom for the exact definition of the research process. Due to this, an approach described by Vaishnavi & Kuechler is used (Vaishnavi & Kuechler, 2015). It clarifies the outcomes of the design cycle and focuses on defining methods and evaluation criteria for created artifacts. This applied methodology consist of five phases: *Awareness of Problem*, *Suggestion*, *Development*, *Evaluation*, and *Conclusion*.

- Awareness of Problem, described in the motivation section, focused on new challenges in user authentication and fraud detection on mobile platforms for financial services.
- Suggestion, where in this case a potential solution in terms of behavioural biometrics authentication system is designed.
- Development, where findings from the literature review are analysed and lead to the creation of the main research contribution. A method will be designed, which will try to solve the problems presented. Outcomes of the literature review, presenting all the possible advantages, disadvantages and research gaps that need to be studied for the development of the method will influence the development phase.

- Evaluation phase and Conclusion phase summarize the results of the study. Main part of the first phase is that the artefacts designed are evaluated based on the defined criteria and validated with the requirements set by the environment. In the last phase the discussion of the results importance is presented.

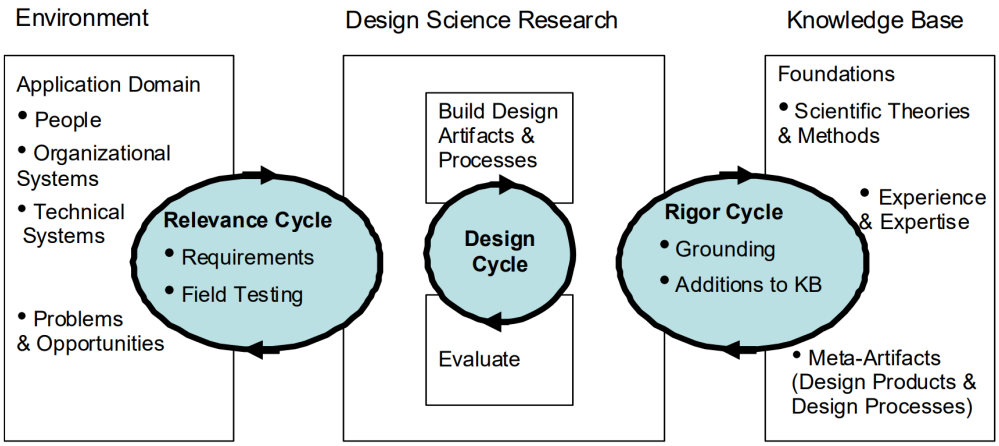


Figure 1.3. Design Science Research knowledge contribution framework. Source: (Hevner, 2007)

Based on the work of Prat et al. (Prat et al., 2014), where authors propose a hierarchy of evaluation criteria for IS artifacts, this work also considers instantiating abstract artifacts as a way of evaluating them. Due to the extensive requirements, both the design of a working artifact meeting the criteria, and further evaluation of its performance will be considered validated contributions to KB - as previously no known work in terms of financial system-ready mobile behavioural authentication solutions was proposed. This work provides an artifact in terms of an authentication method, process and instantiates it, while evaluating and validating its compliance with requirements criteria set by the environment.

1.3.1 Research process

Inline with the methodology used, this dissertation relies on a cyclic process for developing a solution to the aforementioned research. Due to the limitations in time and volume of this work[9], some of the previously done work will only be shortly summarized with references and not described in detail. This previous research is described in the form of cycles (Hevner et al., 2004) presented below. The goal is to present the evolution of research questions, scope of

[9]For the sake of readability, concise definition of research goal and clear evaluation process description.

the work and researched methods, which aimed at achieving verifiable and validable results presented in the dissertation. The publications presented offer more in depth descriptions of the underlying issues connected with behavioural biometrics methods: validation procedures, identity management and adoption of biometrics in financial environment. These research cycles were required, as the topic of behavioural authentication was very broad and multiple literature reviews and experiments were needed to specify requirements that support development of a relevant contribution to the knowledge base (KB) as a result. The previous work covered the following topics:

- **First cycle** - spanning end of 2017 to 2018, fully finished and included all three elements from the knowledge contribution framework presented in the Figure 1.3. Its goal was to perform a preliminary analysis on the feasibility of using behavioural biometrics for anomaly detection and the issue verification of the approach results. Two early papers (Kałużny, 2017; Kałużny & Filipowska, 2018), with one being featured in the Journal of Universal Computer Science, were an attempt to utilize behavioural profiling methods as a reliable authentication and anomaly detection method. The first paper provided a description of methods in the literature and verified their applicability in a potential anomaly detection scenarios, along with the detailed comparison of approaches. After SOTA (state of the art) review the relevance cycle allowed the specification of the requirements for an anomaly detection method utilizing behavioural profiling. The second paper, based on an approach similar to early card fraud detection systems implemented in banking environment, developed a new method for detecting anomalies and providing authentication based on geographical features. The method included novel approach to the detection of geographical anomalies in different dimensions (time, location, sequence). The approach presented met the requirements for creating an authentication method and through the design cycle the results were validated in multiple scenarios on a large dataset. Obtained error rates were promising, with the method providing about 4,5% EER on sparse datasets. Unfortunately the method developed was not directly applicable for financial services authentication scenario due to privacy issues observed. What was the main contribution of this research to the KB and the result of employed rigor cycle were additions to the evaluation and validation approaches for behavioural biometrics methods. Novel evaluation methodology was presented that emphasized the role that the validation approach design plays in influencing the resulting

error rates achieved. This finished the first research cycle, answering all of the preliminary research questions. As a result of those contributions, the overall usefulness of behavioural biometrics methods for authentication was proved, but further work was required to find methods better suited for the environment of financial services.

- **Second cycle** - during 2019 and early 2020 a second iterative cycle took place, focusing on finding the methods suitable for the financial environment and continuing the SOTA review and the relevance cycle in the domain chosen. As the previously chosen methods were not suitable for the chosen domain in terms of their characteristics, an answer to the above mentioned issue two papers were written (Kałużny, 2019a; Kałużny & Stolarski, 2019). Their contribution mostly focused on defining the method's requirements and finding applications and use cases for behavioural biometrics methods in the financial domain. By designing potential scenarios of use, contribution of those two articles resulted in significantly more precise requirements for the authentication method in the domain chosen. In-depth analysis of the privacy threats, length of the authentication process, ease of use and user acceptability of physical and behavioural biometrics, along with the literature review, focused on the details of those methods and gave partial answers to the RQ1. These results however were not formalized enough to create a requirements model but served as a basis for further research. The family of touchscreen behavioural methods was found to initially fit the requirements and as a result the detailed SOTA of those methods was studied in the next article (Kałużny, 2019b). Parts of this analysis, directly relevant to the dissertation are summarized in Chapter 3.4. As an effect of this analysis, the research gaps in the literature directly connected to the RQ2 were identified. Researching those issues resulted in the initial tests of touchscreen based methods on Touchalytics dataset and three separate student projects in which author's supervision or co-supervision with the co-promotor was employed. The effects of those projects included: developing a Proof of Concept Android touchscreen events capturing application by Michał Skrzypek, preliminary analysis of user behaviour differences on BrainRun datset by Aleksander Anku, and experiments on UI design and mobile application infrastructure resulting in an application that captured user's touch behaviour by Paweł Wojciechowski & Weronika Wąsowska. The last project provided an analysis of touchscreen UI's in modern applications that studied the connections between the common interface elements and touchscreen actions they can capture. The designed appli-

cation from the last experiment was also used to collect a dataset of touch behaviour (referred to later as Own Dataset) that is used for the method's evaluation and validation in this work. Unfortunately the results of the second cycle were not broad enough to provide significant additions to the KB and due to that a third cycle was required. It focused solely on the use of mobile touchscreen behavioural methods potential extensions to the approaches employed by the literature.

The outcomes presented in this dissertations are the results of the third iterative cycle. Inline with the descriptions of research goals presented in the Table 1.2, the process began once again from the relevance cycle and requirements engineering presented in Chapter 2, to the design cycle in Chapter 4, along with the validation of designed artifact - mobile financial services authentication method and the description of KB additions in Chapter 6.

This work is also connected to the notion of identity management, by a proxy of studying authentication processes and collecting or storing phone generated data to manage a user profile/pattern. The analysis of those two topics was enriched by conclusions from an additional co-authored publication (Szczekocka et al., 2016). Summarizing, overall this dissertation is supported by six different publications, in five of them the doctoral candidate is either the first or the only author.

1.3.2 Artifacts description

The first and main guideline for the DSR focuses on producing a identifiable and viable design artifact (Hevner, 2007). This artifact must correspond to the relevant and important problem in the research field - which has been identified as a requirement for an authentication method in mobile financial environment. However, despite the applicable nature of the results, the methodology assumes different types of artifacts. Constructs introduce terms and precise vocabulary used to describe the problems and solutions. While some of the artifacts may have been the results of previous work, aiming at expansion or unification of the constructs in the literature is a contribution to the theoretical constructs. This can be tied to the RQ1 and RQ2 in this work, and the precise vocabulary is also presented in the glossary of this work. Models use constructs to create the representation of problems and potential solutions. Methods are tied to a specific solution to the aforementioned problem and may represent algorithms or text descriptions. Instantiations provide proof that the constructs, models and methods designed

can be used in working systems and prove their feasibility. They may be a successful implementation of a method on real data that follows a rigorous evaluation process which measures the utility, quality, and efficacy of the design. In the dissertation all of the artifact categories mentioned are presented, and are listed in detail in the Table 1.3.

Table 1.3. Description of the artifacts designed and presented in this dissertation.

Artifacts	Description
Constructs	The unification of used concepts is provided in the glossary A. The respective financial concepts are described in Chapter 2. The description of concepts connected to authentication and behavioural biometrics is presented in Chapter 3. Building the model of the requirements, the concepts used in it are described in Chapter 2.
Models	Based on the concepts introduced, a model of requirements for authentication methods in financial environments is presented in Chapter 2 together with approaches to the behavioural authentication included in Chapter 3.
Methods	Chapter 4 presents the authentication method designed to solve the research problem stated in the dissertation. The use of specific features, extracted from user behaviour provides a formalization of the presented design.
Instantiations	In Chapter 5 the method is evaluated on multiple datasets what presents its compliance with the requirements and the feasibility of implementation. Further on in the chapter an approach utilizing an implemented model is validated for financial use case scenarios.

Source: own elaboration

The artifact designed in this work must comply with the criteria in which it is evaluated and validated. The choice of the criteria must be connected with the requirements, which are directly tied to the environment of the application domain mentioned before. Providing the exhaustive list of possible criteria is impossible, due to the varying nature of the designed artifact. However, an approach presented by Prat et al. (Prat et al., 2014) (visible in the Figure 1.4) may be utilized to pinpoint criteria which are especially important for this work. The model divides the evaluation into 5 distinctive areas:

- Goal - which captures the degree to which an artifact produces its desired effect in solving the research problem. The generality of the approach is also described, to define the scope to which the results are applicable and significance of the designed artifact to the domain that was chosen.
- Environment - should verify consistency of the artifact with people, organization and technology. For example utility, common to people and organization, measures the qual-

ity of the artifact in practical use, while ease of use may be translated into the usability of the solution. From the technological standpoint the method should utilize recent technologies to provide an advantage in terms of its utility, but must remain consistent with the existing infrastructure of the organization.

- Structure - may be used to evaluate the models and concepts introduced in the work. Present the completeness of researched problems, while retaining the clarity of researched results and certain level of consistency and homomorphism with similar models in the literature.
- Activity - is measured on the dynamic aspects of the artifact. It may measure the method behaviour and is established as a demonstrated agreement with the results of existing experiments (Prat et al., 2014).
- Evolution - quantifies the ability for adaptation and capability to remain functional despite the fluctuations of the environment.

From those areas, a list of evaluation criteria was chosen for this work. It is presented in the Table 1.4. While the goal criteria will mostly be tied with the research goals and the thesis, environmental dimensions are covered by Chapter 2. As the main contribution of research presented is to develop an authentication method, its innate criteria for performance, completeness, consistency, accuracy, performance and efficiency will be used.

1.3.3 Reseach contribution classified by Hevner et al.

According to the classification of knowledge contribution used by Design Science Research (DSR) (Gregor & Hevner, 2013), the contribution achieved may be classified depending on the maturity of the solution and application domain, as shown in the Figure 1.5. In terms of application domain it can be determined by the focus of current research in the topic area (understood here as a domain). Areas with low maturity tend to have different approaches to the same problem with varying results, no baseline for methods or experiments and testing methodologies and processes which are more or less unpredictable. As it increases, the clear definitions, frameworks, guidelines and benchmarks are created, up to the point where process is controlled, its drawbacks and pros are nearly exhausted by the literature created and current research focuses on achieving higher results on already existing benchmarks. For assessing the maturity of the solution (method) one should follow the same principles, but the focus is on the method/artifact used. For example, the use of novel machine learning techniques in eco-

Table 1.4. Evaluation criteria for the dissertation artifacts.

System dimension by Pratt et al.	Evaluation criteria	Measurements	Dissertation reference	Corresponding goal
Goal	efficacy	Meeting the thesis assumptions, results of evaluation and validation processes.	Chapter 5	Thesis
	validity			
	generality			
Environment	[C1,A2] Consistency with people	Consistency with the requirements model for the method.	Requirements model, Chapter 2, Section 2.3	RG1
	[B1,B3,A2] Consistency with organization			
	[B2,P1] Consistency with technology			
Structure	Completeness, homomorphism	Description of multiple behavioural biometrics methods characteristics and comparison of achieved results. Use of existing literature for classification of methods in the SOTA review.	SoTA review in Chapter 3	RG2
Activity	Completeness	Functionality of the method and feasibility of implementation.	Validation in Chapter 5	RG5
	Consistency	Comparison of results on different datasets.	Verification in Chapter 5	RG3, RG4
	[A1] Accuracy and performance	Accuracy and EER (Equal Error Rate) measurements.	Verification in Chapter 5	RG4
Evolution	Evolution, learning capability	Ability to extend the method for new features and dynamically retrain the underlying model with new data.	Validation in Chapter 5	RG5

Source: own elaboration based on (Prat et al., 2014) model

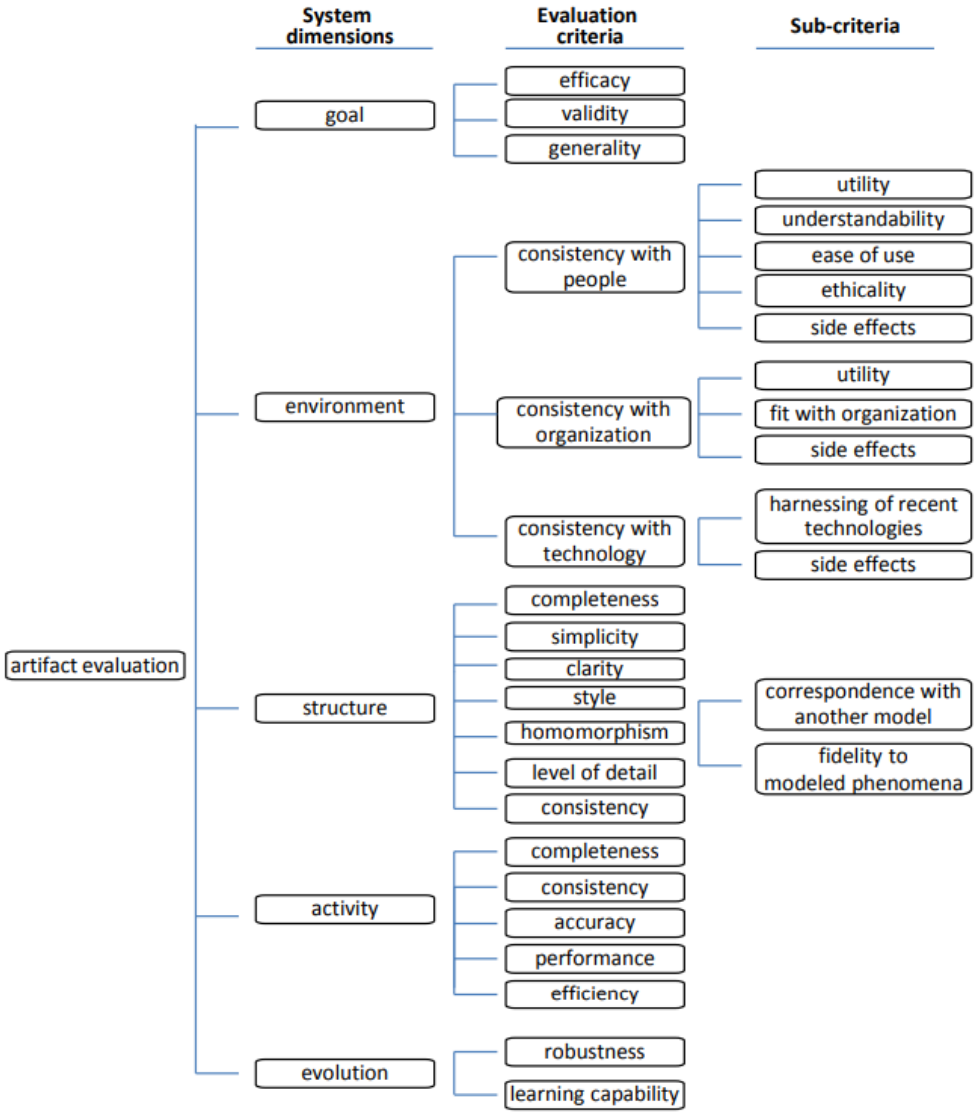


Figure 1.4. Criteria for the evaluation of DSR artifacts. Source: (Prat et al., 2014)

nomics could be considered exaptation, as the method is widely known but it could be used to new problem in the domain. The interdisciplinarity of this approach requires one to define (base) field from which the method is transferred. The clear definition of contribution is not always an obvious task in those situations.

In the case of this dissertation, from the perspective of economics and finance, behavioural biometrics methods are transferred from the field of Computer Science (CS) to be applied in the scenario of mobile financial authentication. The methods have appeared in CS field as early as 1989 (Buschkes et al., 1998) for detecting anomalies in telecommunication networks, with a detailed description of potential approaches in 2008 (Yampolskiy & Govindaraju, 2008). But this is a very broad field, and the topic of this thesis is limited to smartphones and focuses on

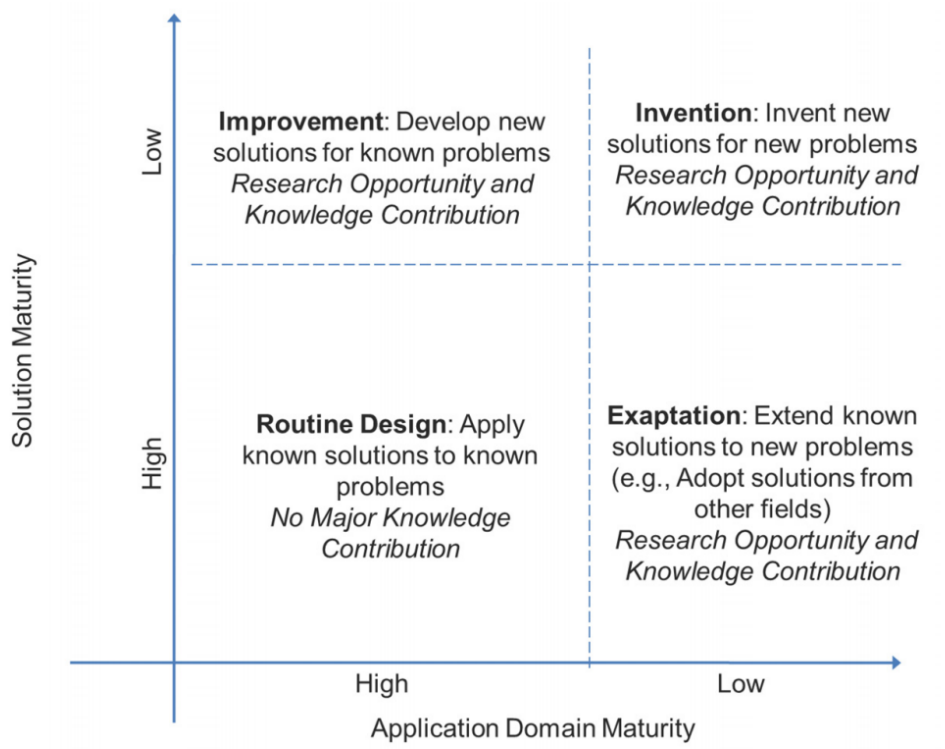


Figure 1.5. Design Science Research knowledge contribution framework. Source: (Gregor & Hevner, 2013)

methods meeting the requirements criteria. Also, not all methods have such a long history - touchscreen authentication appeared only in 2012 (Frank et al., 2012) and the creation of evaluation frameworks, definitive baselines for methods performance and generality of achieved results are still up for discussion in this field[10]. Overall, analysing large-scale, automatically generated behaviour from mobile devices in search of profiles and anomalies is definitely not a widely known topic of research in economics and finance domain. Due to this fact this work can be considered as exaptation, as significant changes and improvements need to be applied for transferred methods to extend them to new problems and the environment of mobile financial services. While this work contribution is clearly aimed at the domain of economics and finance research, this does not exclude that by the development of said method the knowledge base of Computer Science may also be enriched. Further study of those methods in a specific domain and their evaluation results may contribute to both fields of research. Due to those premises, this work may be considered having an interdisciplinary character. It includes a technical part for creating the method which utilizes algorithms and apparatus deriving from Computer Science research, while also containing a requirements analysis and validation, which provide con-

[10]This is described in depth in Chapter 3.

tribution to economics and finance research in terms of use of behavioural biometrics methods in this domain. This is possible mainly due to the DSR goal oriented nature of research and the methodology especially suited for this type of contribution.

1.3.4 Research contribution classified by the field of research

As stated above, the contribution of this dissertation is aimed at the domain of economics and finance research. To prove the validity of this statement, short description is provided below, which is also mentioned briefly in the motivation 1.1 and expanded in depth in Chapter 2, especially in the Section 2.1 dedicated to the role of innovations in financial services.

The dissertation is centered around the problem of authentication and frauds in mobile financial services and the issues of current authentication methods used in this domain. From the perspective of economics and finance studies this dissertation covers the topic of developing a financial innovation, categorized as process (in terms of fraud detection systems and authentication) and partially also a product (by developing a prototype of higher usability authentication processes for mobile financial domain applications). Both of these are types of innovations which are an inseparable part of finance research (Marcinkowska, 2012). Technical innovations (such as authentication methods) are a part of financial innovations, which are a key competitive factor on the financial market. The artifacts that are proposed in this dissertation influence the financial services sector by providing new and improved services. These proposals are evaluated and validated, according to the research rigour required by the methodology chosen in the dissertation. Hence, the findings of this work, supported by computer science algorithms such as machine learning method, and are by an extension part of widely understood financial innovations in the same manner as new financial instruments utilizing mathematics or organizational processes applied from management science are, when applied to the financial domain.

The works also builds upon the notion of behavioural economics, as it tries co capture and measure user behaviour in mobile financial applications. Instead of relying on traditionally used statistical profiles for fraud detection it utilizes behavioural economic approach of learning how user behaves and what behaviour may be normal or abnormal (in case of theft or fraud) for a particular user. This work uses machine learning algorithms such as Support Vector Machines (SVM), Random Forests (RF) and gradient boosted versions of those (XGBoost) to build a model unique for a specific user. This combination of an economic research carried out by methods

originating from the field of computer science have been a trend in the recent studies from the field of economics and finance research and are often applied with great success.

1.4 Structure of the dissertation

The dissertation consists of 6 chapters, including the introduction and summary. Rest of the chapters can be divided in three groups: focused literature analysis, design and verification and validation.

First part of the dissertation, consisting of Chapter 2 and 3 and the motivation section of the introduction presents the results of the literature analysis. Firstly, the domain of the problem is described in Section 1.1. The issues of mobile financial applications are then expanded upon in Chapter 2. This chapter not only analyses the potential changes that are tied with the transformation of financial sector to a more mobile-centric model, but also discusses the role and need for new authentication methods in this environment. The chapter lists the examples of successful innovations like BLIK or biometric authentication in mobile applications and their influence on the role of banks and financial institutions compared to their potential fin tech competitors. As a result of the analysis of customers and banks needs, a model of requirements for the mobile authentication method is specified at the end of the chapter.

Chapter 3 presents the types of biometrics and the authentication process itself. Besides describing the characteristics which are tied to the biometric traits that are used in authentication, the chapter discusses which of the traditionally present characteristics can be observed for behavioural biometrics patterns. By comparing the physical and behavioural biometrics the chapter presents advantages and drawbacks of both. Further on a detailed analysis of behavioural methods is presented along with the state of the art analysis, which leads to the comparison of methods that aims to choose the best method for authentication in the mobile financial application scenario.

Second part of this work, covered by Chapter 4 is focused on designing a behavioural biometrics authentication method relying on touchscreen interaction patterns. The chapter presents the assumptions, design considerations and limitations which are tied to the criteria for the evaluation of the method. Secondly, the chapter in detail describes the design of the method, including the description of features which are extracted from touchscreen patterns

and additional sensors. Finally, the chapter presents the classifiers which have been chosen for the method and the evaluation criteria for their optimization.

Third part of this dissertation, spanning over the 5'th chapter describes the process of evaluation and the validation of the proposed method. First of all, it presents five distinctive experiments on multiple datasets which test the method's performance and evaluate its accuracy and error rates. First three experiments focus on a classification problem of fraud and user activities and point out the discriminatory power of the features and their distribution among observed touch patterns. The third experiment especially points out to the features which are the best in differentiating between user patterns. Next experiment presents the possibilities of expanding the model with tap classification. The last experiment includes another large dataset in addition to the Own Dataset prepared for evaluation, and tries to handle issue of recognizing user gender and age for fraud detection mechanisms, possibly helping in authentication insider threat cases[11]. Finally, after proving the method's performance and error rates are satisfactory for the thesis assumptions in the evaluation process, the validation process for the produced artifact is explained. In this portion of the chapter, we provide scenarios in which the method may be used to improve usability, security and enable novel authentication and authorization processes. Possible communication schemes with fraud detection systems and architectural decisions in terms of data processing are evaluated. Further on the limitations of the evaluation and validation processes are discussed.

Finally, the sixth and the last chapter presents the summary of the work and contains an overview of the most important research results along with the work's contribution and the further work possible in the dissertation topic.

[11]See glossary A.

Chapter 2

Electronic and mobile banking and payment applications

The goal of the chapter is to describe the role of banking and payment systems in the current digital economy, characterizing trends and issues the financial sector is currently facing. As most of the described problems are connected with the conversion to the mobile-centric model, the chapter aims at engineering the requirements for the authentication method for the mobile financial environment. Taking into consideration main issues in the adoption of mobile-centric model of services it presents and classifies the requirements for an authentication method which may solve some of the issues presented in the chapter. This in turn provides results for the **RG1**, creating a requirements model fit for the financial environment - compliant with the technological, organizational and customer's requirements and designed for the mobile authentication process environment itself.

The structure of the chapter is as follows. First part of the chapter emphasizes on the role of innovation in financial institutions from both the business and scientific perspective placing it as a valid base for research and a crucial business requirement. Later part is focused on characterizing the market and trends shaping the services and products offered in the sector, pointing to the mobile applications as a defining trend influencing the development of new services. As developing new and improved services is necessary for the sector, current trends shaping the services and products offered in the sector are described. Focusing on the acceptance of mobile-centric financial services, the chapter presents the problems tied with authentication process in mobile financial applications from the perspectives of customers and financial sector representatives. It thoroughly presents the current issues that the financial sector faces in this

regard connected with the technological and legal requirements e.g., mobile banking, PSD-2 directive, Open Banking architecture and fraud detection in this environment. At the end of the chapter a structured model of requirements for new authentication solutions in the above mentioned environment is provided.

2.1 Role of innovation in financial services

Financial innovations are very specific - as they are rather short-lived (impermanent), easy to copy, burdened with high risk and often do not lead to achieving competitive advantage (Marcinkowska, 2012). This may in turn raise a question if financial institutions even need to pursue innovative services. On one hand financial innovations both provide mechanism to finance innovative technological projects with high investment risk, but technological and economic progress result in the higher complexity of business processes and new types of risk in financial systems. The diffusion of innovation is high in this domain (Błach, 2011), as solutions can be easily tested on non-domestic markets and can easily be adopted by the competitors. This in turn forces the financial system and financial markets to adapt to the changes, to be modernized according to the new requirements of the business entities and to the challenges of the modern world.

Another reason justifying innovation may be that nowadays banks and financial institutions need to develop new roles, better justifying their place in the economy and society. The underlying digital and mobile transformations in financial sector causes disintermediation - where IT allows multiple transactions without the traditional role of banks as brokers. With that in mind, the threat from non-sector competition rises. Telecom and IT companies offer more and more services that are substitutes for what traditional financial institutions have to offer. Therefore, despite being burdened with high risk, innovations are essential to achieve success on the market. The work of Marcinkowska (Marcinkowska, 2012) studying the role of financial innovations provides a good methodological review of their importance in the research connected with the finance domain. The determinants of such innovations may be classified as:

- Market - changes in the financial environment. Mostly caused by the cost of providing services and directly influenced by regulations, transaction costs, information asymmetry (in terms of risk-management) and competitors (also multi-sector ones like telecommunication companies employing financial services or FinTech). It can be said that **these**

are the requirements of banks and financial institutions as actors, as they most closely resemble free market tendencies of every business entity for providing competitive services.

- Demand based - connected with customer requirements, not only for specific services but also segments, channels and infrastructural demands. The needs of customers are increasingly important in the sector, as their requirements grow along with their bargaining power caused by more freedom in choosing their provider in the sector (including entirely new companies deriving from FinTech which offer highly competitive services). As such, this determinant is shaped by the **requirements of bank's consumers.**
- Regulations - laws and requirements concerning the institutions for example regarding the capital structure. Those can also include regulations which highly influence the Infrastructure internal processes such as PSD 2 directive (Europen Comission, 2015) or GPDR. As such this determinant **shapes the requirements deriving from regulations - nowadays mostly concerning risk, privacy and data protection.**
- Technology - the development of technology is the most visible drive of innovations in financial sector. Along with the possibility of processing new sources of information and improving the means of processing and communication, the technology highly influences instruments (allowing e.g., dynamic pricing), markets (cryptocurrencies, digital goods), institutions (FinTech), Infrastructure (BLIK) and norms (e.g., cryptocurrencies regulations).

The types of those innovations may be distinguished as:

- Product - providing new product, or improving the services provided in terms of: technical specifications, components and materials, software, ease of use and usability, or new functionalities. The main drivers of those innovations are technological progress and achieving new clients.
- Process - understood as improving both internal and external processes. New service strategies and plans, new interfaces and systems for delivering services (e.g., mobile banking, allowing loans in mobile channel), new technological capabilities (customer and product analytics, anti-fraud systems).
- Organizational - including all new management strategies, methods connected with company strategy, workplace policies and relations with the business environment, with goals aimed at improving the financial result of a company.

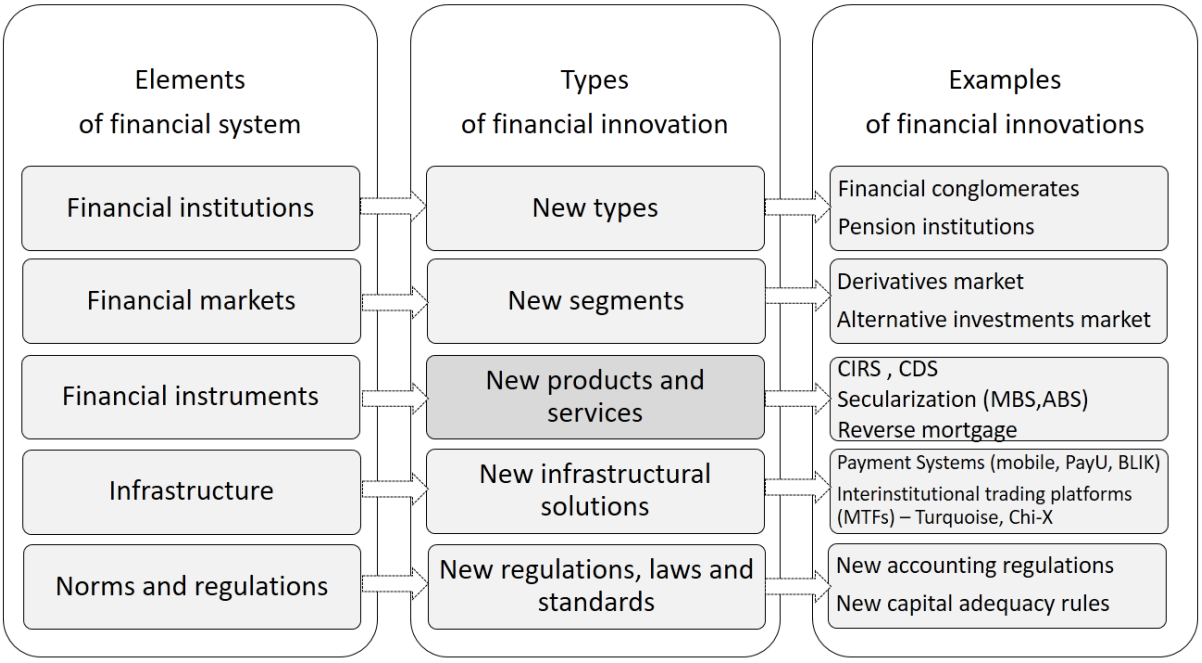


Figure 2.1. Classification of financial innovations in narrow understanding. Source: own development based on (Marcinkowska, 2012)

- Marketing - encompassing all changes in the design, distribution, product promotion and pricing of products.

As shown in the Figure 2.1, categorization of innovations is dependent on the element of the financial system they are supporting. The figure represents the narrow understanding of innovations, where widely known examples are shown. But they do not determine the depth and complexity of today's financial system. **The innovations should not be understood narrowly as only new financial products, but all improvements to those above mentioned categories and solutions influencing all elements of the financial system** - markets, institutions, financial instruments, operations, norms and regulations. As such technical innovations (such as authentication methods) are a part of financial innovations, which are a key competitive factor on financial markets.

Similarly, the type of innovations can be classified by functions they provide as presented in the Figure 2.2. Financial innovations to bring benefits, should decrease the level of risk, close the information gap or lower transaction and tax costs or enhance the existing element's efficiency, stability or usability (Błach, 2011; Marcinkowska, 2012). For this work, the risk management function is especially emphasized due to the applicability of the results in authentication processes, detection of anomalies for anti-fraud systems and use in a transaction level

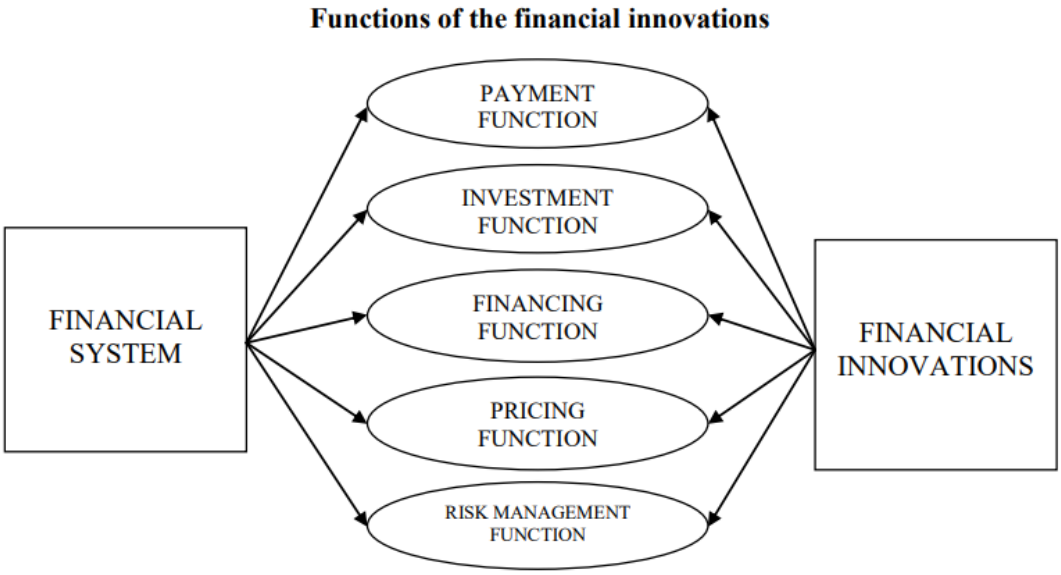


Figure 2.2. Functions of financial innovations. Source: (Błach, 2011)

risk-assessment and authorization. As the innovations are encouraged and potential areas of application are subject to the domain of finance and economics, developing financial innovations is a valid topic from both the research and practical domain perspectives. And while the arguments presented before prove its importance for the sector, what is still needed to be clarified are the most promising areas where these innovations may be applied to have the highest influence on the future of the financial sector.

2.2 Role of banking and payment systems in digital economy

Current changes in the digital economy shift the paradigm of services provided to be more aligned with the current technology. Based on the concept of knowledge based economy (Olszak, 2007), the companies need to utilize the wealth of information available to them to improve their services and internal processes. They also need to use the IT tools which allow processing this information into business applicable knowledge giving the competitive advantage.

Currently the financial market is a very dynamic environment, characterized by a high rate of implementing innovations. Studies showcase that the Polish financial market is one of the most innovative ones (Bolesławski & Nowakowski, 2016; Visa, 2016). Appearance of new services such as virtual credit cards only emphasizes its recent dynamics. It also makes it more similar to IT and high technology markets (Luo et al., 2010). Banks must face competition (including

FinTechs), market uncertainty, manage different risks and also manage the technological difficulties. It is also a market, where the decision to change financial provider may be caused by small singular obstacles caused by minor security incidents, old legacy technology, or even lack of usability. In line with that coming up with innovative services, despite burdened with risk is often the source of the competitive advantage that defines the leaders in the market and may maintain banks strong position compared to their newly appearing competition.

On the other hand, the ubiquity of the mobile phones mentioned in Section 1.1 also influences the whole financial sector. Not only in banking, the whole financial sector uses the mobile and electronic channel more often. But due to the very high dynamic of the mobile channel it is the most important for customizing and improving the services to meet the customers' needs and increase the level of security.

Another challenge is posed by new payment technologies that are being introduced with an ever growing marker share: Apple Pay and Google Pay and external payment applications such as Polish BLIK, PayU or PayPal. Also, the model of payments is changing as smartphones are also used as mobile payment terminals and there is a an emerging trend of paying with the use of smart watches. As more people are using mobile phones, their requirements shape a new model of mobile-centric (Srinivas et al., 2018) in which two main trends are clear:

- **Integration with new technologies** - banks are losing their dominant position in the financial sector. With the appearance of new FinTech companies, multiple third parties handling parts of the payment and account management process (PayU, Przelewy24, PayPal, Blik, GooglePay, ApplePay) and technologically focused competition (Revolut), they must adapt to the new environment or will be marginalized. Innovation has become not only a novelty but a necessity. With new legislative initiatives such as PSD 2 and Open Banking the adoption to the third parties IT infrastructure can be considered a requirement. This in turn means that financial sector institutions which will integrate their infrastructures with the new initiatives in the early stages will gain a competitive advantage either securing or largely improving its position on the market. However, institutions which were traditionally trusted with customer finances and their data must secure them from theft and misuse. This becomes hard due to the technical requirements - as such adoption of security improving technologies is of the most importance. Banks being the institutions trusted the most in the sector may be naturally asked to handle most of the transaction's risk in this new Open Banking infrastructure, which is

challenge but also an opportunity to retain a strong position in this new mobile-centric model (Ficowicz, 2018).

- **Focusing on customer requirements** - as users tend to have significantly more options to choose from, they more vocally express their needs concerning services provided. Mobile-centricity is a natural example of this change, but with it come also various user requirements regarding the security, usability and the quality of the service itself as proven by the Visa and Mastercard studies (Mastercard, 2018; Visa, 2017a, 2017c)[1]. Customers require usable services, and will choose a financial service which is the most useful and still permitted in terms of security - as the risk is mostly transferred to the Bank handling the transaction. This makes increasing the transaction security while not hindering the usability a matter of customer retention which will shape the sector in the upcoming years.

Similar to those above mentioned notions, Capgemini in their analysis (Singh et al., 2020) are confirming these aspects. Focus on the digital channel, Open Banking and analytics with the use of AI and machine learning technologies drive banks to further integrate with Open-Banking and FinTech and RegTech companies. This collaboration results in customer satisfaction improvement but mentions identity and risk management (also in the under-represented areas such as digital channel lending) as one of the key factors shaping the sector nowadays. Currently banks face new potential competitors on the market - FinTech companies. In 2018 it was valued at about 127.66 billion, and is expected to grow to 309.98 billion at an annual growth rate of 24.8% through 2022 (The Business Research Company, 2019). FinTech is currently a very dynamic sector which highly relies upon innovative technologies in the process of providing financial services. It is also often considered an alternative to traditional financial services and banking which encourages the competitiveness in this market. Similarly, the trends for this sector, defined in 2017 by the American Banking Association for the next decade (Morgan, 2017) include: digital lending, biometrics, use of customer data, with RegTech and AI as the tools for it. Developing solutions based on customer data analysis and biometrics with the use of right analytic tools will help in transforming the services and will increase the technological interoperability in the sector. On the other hand the new role of banking institutions as identity and transaction risk holders and shift of the payment systems to the mobile and digital channel showcases the importance of the dissertation topic to the current trends

[1]They are described in Sections 2.2 and 2.3.2.

in the financial environment. This also proves that the development of new mobile channel suited technologies which can improve the security and handle transaction risk is of the the utmost importance from both the customers' and service provider's perspectives. The only issue left to answer is if large institutions should be encouraged to pursue such innovative services and if constant development of new services is better than relying on the security of known mechanisms.

2.2.1 Mobile payments and banking market

From the research report conducted by VISA - Digital Payments Study 2017 (Visa, 2017b)[2]. 77% of European citizens utilize smartphones for digital payment and banking services. 68% of users use some form of digital wallet or use card-on-file service (where the websites store their payment details). Those statistics appear to be more meaningful in case of younger respondents (aged 18-34), as 86% of the respondents from this consider themselves „Mobile Money users" and 92% expect to be considered among this group within three years. The findings for the EU market are similar to the results achieved for the Polish market, with a slightly higher dynamics observed for mobile banking applications use in Poland - 67% of customers compared to the 62% for European average (Visa, 2017c).

The group of mobile money users is increasing[3]. Currently nearly 50% of the customers utilize mobile money applications for everyday shopping, according to the Mastercard study from 2016. Similarly, in 2021 (Strohm, 2021) three in four Americans (76%) used their bank's mobile app within the last year for everyday banking tasks.

According to the broad Mastercard study, the share of customers using mobile banking applications in the 4th quarter of 2016 was 55% among all the customers utilizing digital banking (for example also using the bank website). This makes up for 18% of all banking customers in Poland (see Figure 2.3), but is also a number that is subject to the significant growth in years. According to the prnews.pl report, more than 16,7 billion Polish users utilize mobile banking (Boczoń, 2020), 8,4 billion are mobile only, meaning they access banking services solely from

[2]Visa commissioned the Digital Payments research with Populus. The research was conducted between June and July 2017 in 22 European countries: Austria, Belgium, Bulgaria, Czech Republic, Denmark, Finland, France, Germany, Greece, Ireland, Israel, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, Turkey and the UK. The total sample size was 42,308 consumers, with approximately 2,000 respondents per country.

[3]In the VISA study mobile money refers to the usage of a smartphone, tablet or wearable to manage money or make a payment in person, online or in-app.

their smartphones. Compared to the previous studies in Poland, we can see that the number of overall customers in this channel doubled since 2016 (Sudoł & Woszczyński, 2018).

Comparing the dynamics of mobile banking and payment applications market, the results observed on the Polish market showcase higher adoption of those services in Poland than in other European countries. For example, comparing the Polish mobile banking user's share of all banking customers (from Mastercard's study), with the PWC study for Germany, we can see over 5 percentage points of advantage in Poland, over the 13% achieved in the study conducted by PWC for Germany in the year 2017 (PwC, 2017). The study also confirms the possibility of significant growth in the future, as 42% of respondents consider themselves willing to use mobile payment services. Overall, the constant growth of digital and mobile channel customers points out to it being the main focus of the current financial services market.

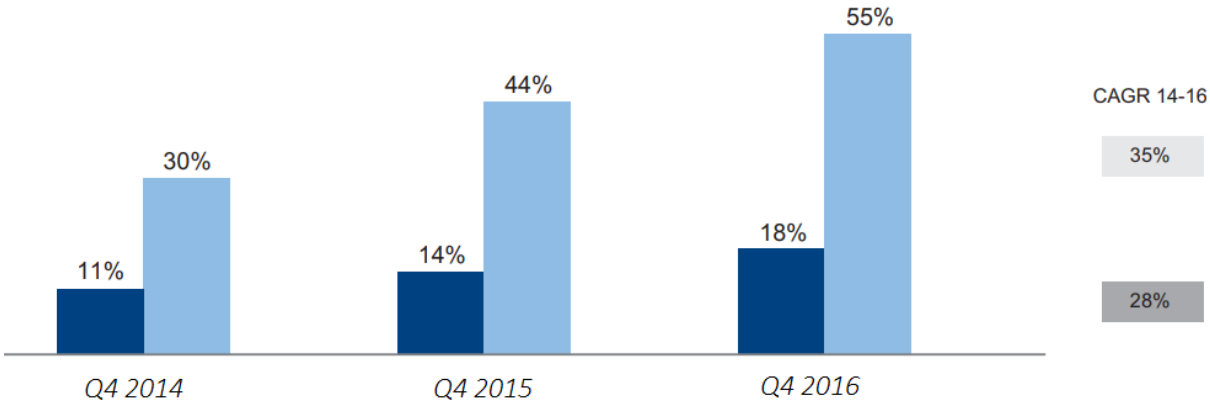


Figure 2.3. Percentages of mobile banking customers among all electronic banking customers (blue) and all customers (dark blue) in the 4th quarters of 2014, 15 and 16. Source: (Mastercard, 2017)

Mobile banking has a benefit of an easier access to user data, which in turn allows for an easier fraud detection compared to the other parts of the digital channel. The customer's conviction about the efficient inner workings of security procedures are key elements which influence customer's choice and prevent churn (Luo et al., 2010). The customers however use their mobile phones not only for secure financial services, utilizing it for mailing, social media and games - the variety of installed applications makes this channel especially vulnerable to malware.

Nowadays more and more providers offer "virtual" or "mobile" credit cards (examples, offered e.g., by Santader Bank are shown in the Figure 2.4). These new types of services nowadays do not require any physical token, RFID or card but are instead provided by the phone itself.

Due to that, the card holder is no longer required to own a physical card as all of the transactions can be provided through phone registered virtual cards. They provide NFC payment services, and are tied to user's financial applications. The popularity of these new services is in line with the mobile-centric trend of the sector, but they rely only on the security of the device and bank's mobile application. This in turn makes their security even more of a priority in modern financial services.

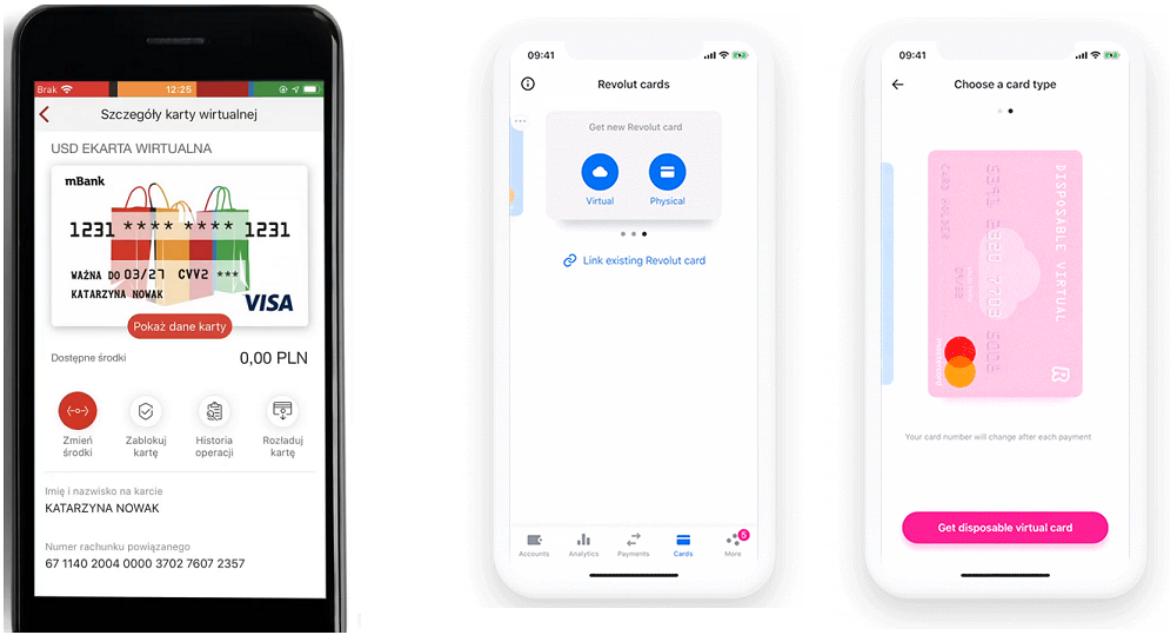


Figure 2.4. Mobile virtual debit cards by mBank (left) and Revolut (middle and right). Source: (MBank, 2020; Revolut, 2021)

2.2.2 New financial services - BLIK

Development of new technologies creates new opportunities for making payments in mobile systems, which also need to consider security restrictions. The example of that is the use of NFC (Near Field Communication) technology - today replaced by HCE technology. Both of them allow using a smartphone in the same way, as a traditional payment card, by storing its virtual representation on the device. Much of the "wallet" providers such as GooglePay, ApplePay and PeoPay use this technology to provide their services.

A different answer to the same issue - BLIK system, developed in Poland, allows using mobile devices for mobile payments and even ATM withdrawal in a varied bank-independent environment. Payments rely on one time passcodes, available directly in the application, which are then passed to the service in which the payment is made and then confirmed by the user

on original mobile device. The overall process is presented in the Figure 2.5, the steps for the procedure beginning from the left. As the service requesting the authentication can vary (ATM, shop, FinTech) the platform is independent and working as a third party authentication provider with the BLIK token generator integrated with the banking application. It however requires authentication in the financial service provider itself - meaning user do interact and need to login to their respective bank (at the step 3 from the left in the figure) to generate a 1 minute valid temporary code. Then, they are often asked to confirm the transaction independently back in the app at the end of the process. This means users need to conduct at least 3 steps in the banking app - login (step 3), generate the BLIK code (step 4) and confirm the transaction (step 6).

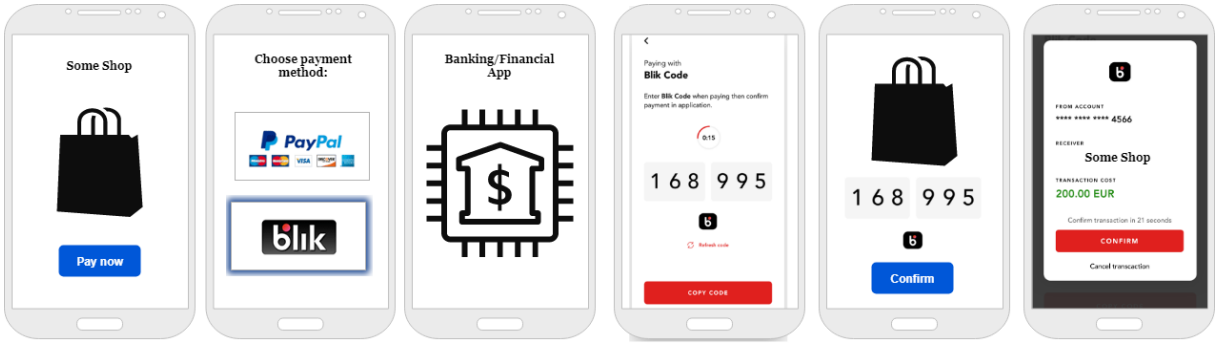


Figure 2.5. BLIK system platform-independent payment process. Source: own development based on the banking app design from (Gadzina, 2020)

.

The BLIK service had become a very popular example of a new financial service in Poland. Initially in 2017, it was used by 7 Polish banks, over 5,3 million users and more than 75 thousands of shops. Today BLIK is actively used by over 5 million people at the end of Q2 2020 (www.blik.com, 2020). This is 67% more than a year earlier and nearly all major banks support it. It is a service with ever-increasing dynamics, as can be seen in the Figure 2.6. The overall value of transactions increased 6-fold from Q3 2017 to Q3 2019, up to 7,6 billion PLN in the latter date. With its highly increasing popularity, the service that appeared in the 2015 as a payment system being a result of the cooperation of multiple banks, is an example of a mobile-centric innovation that has revolutionized the market in Poland. Designed to work in digital and mobile channels, with security of the transaction and ease of use in mind (data for the transfer is passed directly from the payment request party) it has achieved a great success. It also has a benefit of solidifying bank's advantage in an area that was earlier taken over by

PayU[4] and similar payment services all over the world. It is still highly competitive, even with the appearance of Apple and Google Pay services and varying competitors. Summarizing, BLIK is a mobile centered solution that improved the security and usability of the transaction. It did so by developing mobile and digital channels and achieved a tremendous success. This points out that researching similar trends that may improve the payment system in the future is a topic worth of both scientific and market research.

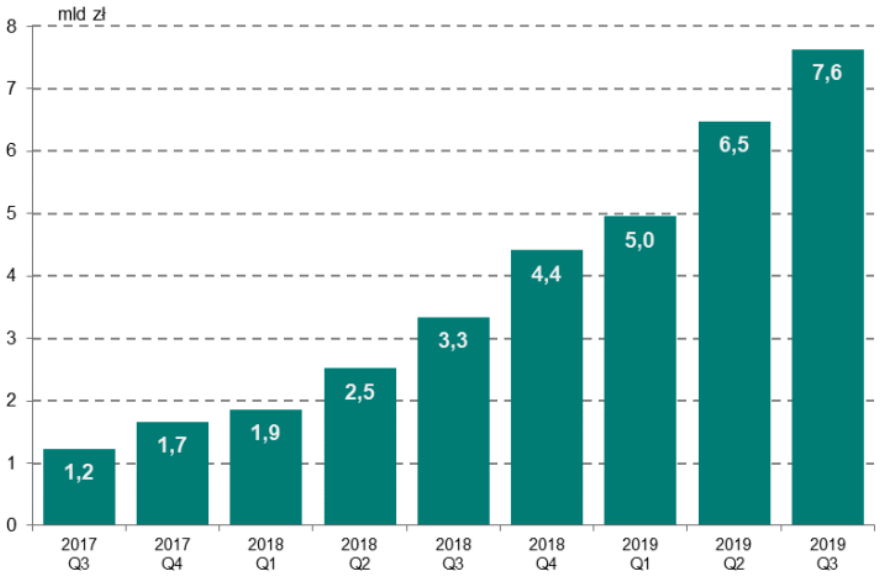


Figure 2.6. Value of transactions in BLIK system in Poland from Q3 2017 to Q3 2019 in billions of PLN. Source: (Narodowy Bank Polski, 2020)

2.2.3 Financial services environment - Open Banking

Based on some of the concepts which may shape banks strategies for the next few years, banks are still seen as a platform that has customers' trust (Świeszczak, 2017; Zakonnik & Czerwonka, 2014), compared to other parties which may hold user data and private information. For example, 63% respondents of the VISA study (Visa, 2017c) have no issues with sharing their biometric details with banks. On the other hand, 42% of people wouldn't want to share their data with social networks, and 84% expressed discomfort about sharing sensitive personal details with social media, including two-thirds (66 percent) who say they are uncomfortable sharing their bank account or payment card details with social networks. Building upon this idea, banking institutions may be nowadays placed as providers of identity management services. Due to the

[4] https://payu.com/

fact that customer trust plays a great role in perceiving banks we can point it as a significant advantage over FinTech service providers. Banks as institutions of public trust are perceived by their customers much more positively in terms of organization and stability that new companies appearing in the financial market. This confidence is much higher (Świeszczak, 2017) for the traditional financial institutions because they are trusted by the public. They may be a natural choice as handlers of authentication and identity management services, while also storing the data required for authentication. Along with those facts, the current role of banks is to provide infrastructure and methods that not only provide safe and secure services, but also can manage user identity data (which may include biometric patterns). This new role as an identity provider is connected with the **Open Banking** initiative that banks are required to comply with.

According to the PSD 2 directive (Europen Comission, 2015) **banks must allow access to their core services through the application programming interfaces (API's) for external providers**. This standard has been called Open Banking, and the general infrastructure is shown in the Figure 2.7. When a customer is using a third party application (e.g., payment website with BLIK code), the service provider is essentially accessing the internal bank architecture by the proxy of an API. This regulation induced type of innovation significantly changes the bank's position on the market. Third parties by using those APIs can generate new services to the bank's customers. These new rules included are aimed at promoting the development and use of innovative on line and mobile payments through the Open Banking infrastructure. This trend influences financial institutions in three main areas:

- Development of services - with the open API infrastructure the competitiveness of the market rises even more, as all of the sector companies (not only banks) can develop new services, including payment platforms. This leads to the creation of better (more usable or innovative) services.
- Identity management and Authentication platforms - as banks hold their main user information, they are considered to be main authorizing parties in the platform. Due to that, they need proper infrastructure and new solutions for managing their users' data.
- Security and risk - similarly to the above mentioned examples, banks need to think wider than their infrastructure, about the security of their user patterns (which they can use in other services) and also provide state of the art risk analysis for transaction authorization

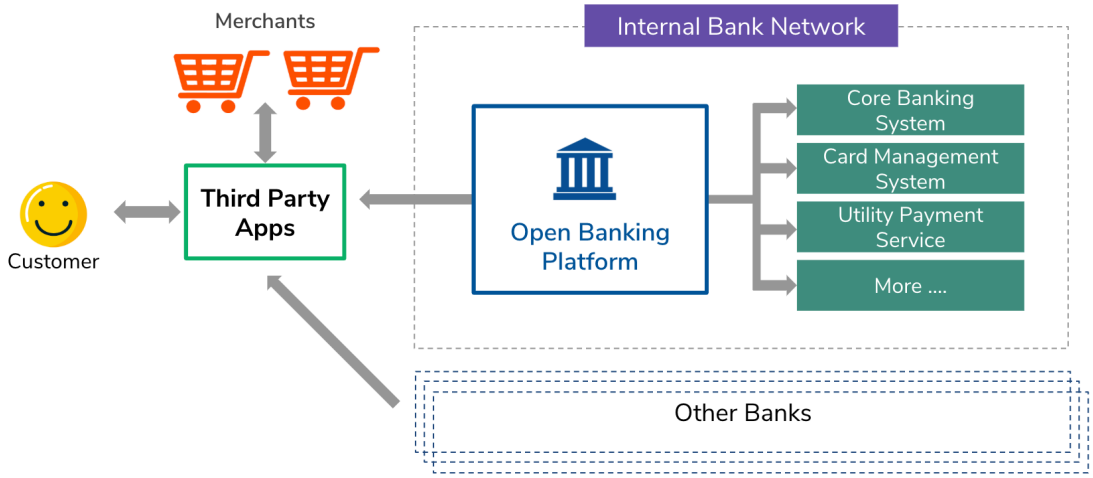


Figure 2.7. Open Banking architecture. Source: (Subasinghe, 2019)

and risk detection. This in turn means the need to use all potential sources of data for assessing risk in real time.

Especially the risk connected with a specific transaction of an authenticated user is an issue worth addressing. Banks need to focus more on the security of the architecture, compared to the third parties providing financial services. Even if they will not be directly responsible for the potential fraud cases, their image and position as institutions of public trust may be hindered in the eyes of their users resulting in the loss of customers.

2.2.4 Fraud detection systems in financial institutions

Frauds, as described in Section 1.1, are not only an ongoing but more of a developing issue in the financial sector. The rapid development of technology and wider access to mobile and electronic banking lacks the standardized and thoroughly tested procedures for confirming user identity. What has worked previously in the banking branch - may not be suitable in the mobile-centric environment. The issue of preventing fraud is making sure that the user identity is confirmed or effectively all credentials are matching, and the transaction is carried out willingly. Today's process of authenticating a user is vulnerable to multiple types of potential fraud and unauthorized transactions cases:

- authenticating a transaction by stolen credentials and/or the users mobile phone,

- unauthorized use of user's account by a person known by him/her (spouse/child) called an authentication insider threat[5],
- unauthorized transaction done by malicious code executed on the application,
- use of a fake application to confirm a fraudulent transaction.

Some of those issues are becoming especially prevalent in the Open Banking architecture, as the risk is mostly borne by banks. Similarly, to prevent potential liability and complications the bank needs to make sure that not only the user's credentials are valid but that he/she is willingly commissioning the transaction. This issue can be named to a wider problem of proof of presence (Samet et al., 2019), described in the biometrics and security domain of Computer Science. To provide a valuable tool to combat current fraud, an authentication factor (possibly inherent to the user) that is hard to spoof or capture is needed. If the pattern is easy to read and permanent[6] it may be easily spoof or captured by a perpetrator. The use of inherent biological of behavioural factors however effectively protects the user from insider threat, as obtaining e.g., fingerprint, retina scan or behavioural pattern is hard to have been carried out on an unsuspecting victim. On the other hand, guessing a password, PIN code or lock pattern is on the other end of the scale being very easy to effectively obtain by observing the user. Using the device of a user as a factor protects from standard fraud cases, but not from malware, insider threat and phone theft (or using user's phone when he/she is unaware of it). Due to the above mentioned issues, to combat current cases of frauds in a mobile channel, a proof of presence authentication method, based on a pattern that is hard to spoof or capture/observe is required.

The topic of fraud in financial services is a broad one, and describing every method, approach and data source which can be used is outside of the scope of this work. However, to describe how behavioural biometic methods may improve this environment a quick overview of the current approaches may be necessary. Based on a recent strategy for fraud detection published by Gartner in 2019 (Vignolo, 2019), which also serves as a capability model for the assessment of the interested institutions (as presented in the Figure 2.8), there are multiple different scenarios. Based on the axes of foresight and hindsight - whether we are reacting based more on a defined set of rules or using derived patterns of behaviour which can be updated.

[5]The most popular definition of insider threat used in finance is known to the doctoral candidate, but the issue of unauthorized access has been referred to as such in the literature, for example (Muslukhov et al., 2013). Hence, the author did not change the name of the issue not to create inconsistencies in the analysis.

[6]Which are the traditionally desirable characteristics for biometric features, as described in Section 3.3.

As traditional models which compare data or rules tend to fail on more sophisticated attacks, and their flexibility is limited. As more rules get added, more valid transactions can be considered fraud, with no way to correct them for a particular user. And while modeling in terms of geography and user device is used now, it relies on a simple statistical trends derived from the population. This "one size fits all" methodology is less effective in an increasingly complex fraud landscape. New analytics and data sets need to be utilized (e.g., machine learning, behavioural analytics, and biometrics), as pointed out by Garner. Security and risk management leaders must create a fraud detection landscape that begins to assess risk from when the customer arrives on their digital premises, where the behaviour is analyzed in a continuous pattern. As this notion is considered to be the most fitting for the new environment in which these frauds are happening, methods which can provide financial institutions with behavioural patterns not directly tied to transactions, but rather created by the whole process of authorizing a transaction, are the most valuable. This type of fraud detection is especially important in the case of Open Banking architecture, as potentially certain level of interoperability in the process would allow service providers to supply additional information to the financial institutions handling the transactions. This new model of fraud detection, where cooperation between parties is encouraged, not only aligns with the PSD 2 directive, but may be a deciding factor in the future of financial sector transaction security.

2.2.5 Summary - modern financial sector challenges

Summarizing the description of the current trends shaping the financial sector's development, a few main challenges being faced by the sector were identified.

First of all, its recent digitization and transformation to the mobile model have caused an undeniable level of work in transforming current services to the mobile model. Meaning developments in the area are of the utmost importance from the perspective of customer retention. More and more services appear in the mobile channel and with the introduction of virtual cards and services like BLIK, mobile devices become the center for accessing financial services by customers. Hence, focus on the security of transactions, customer satisfaction and mobile-app oriented solutions is justified.

Secondly, despite its inherent connection with security and thoroughly auditing the technologies used, innovative services are required in the area of finance and banking services. Innovation is not only a choice, but a necessity, especially considering the rising role of cus-

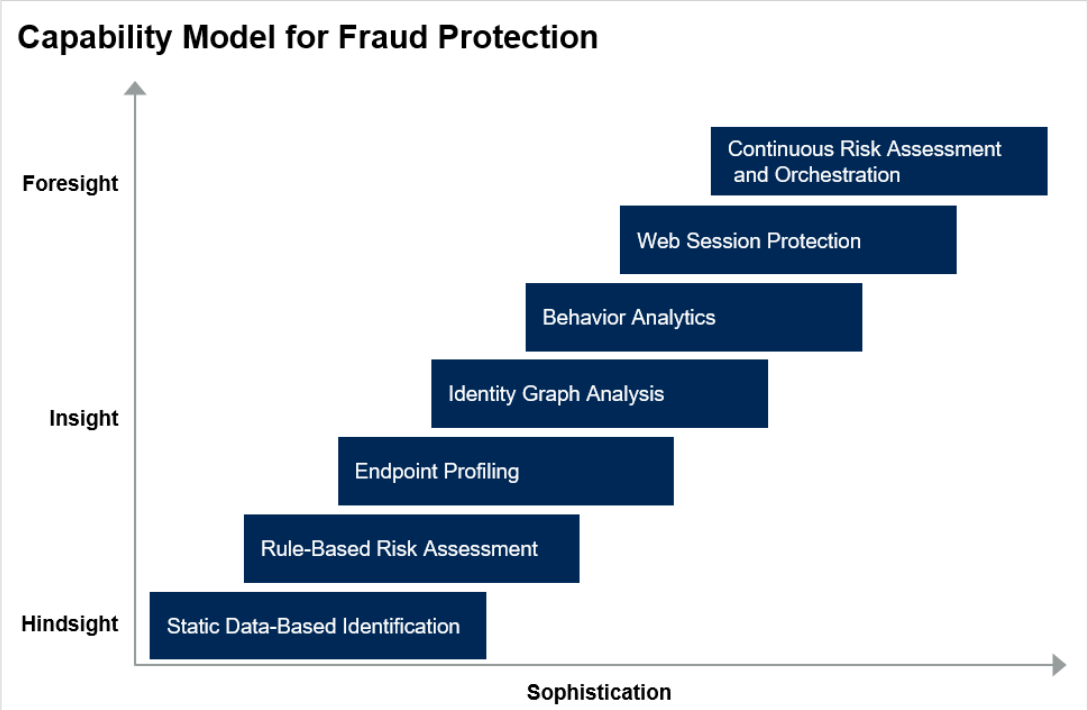


Figure 2.8. Capability model for fraud detection. Source: (Vignolo, 2019)

tomer needs in the new open financial sector. This openness refers to both the increase in competition (in terms of Telecoms and IT giants providing financial services) but also legislative initiatives such as PSD 2 that require financial institutions to open their technical infrastructures for new services.

As this type of openness comes with additional costs in terms of transforming the current architectures, using it to further improve or provide additional service seems to be the rational choice for the large institutions, such as banks. Higher cooperation with innovative services from FinTech area (such as BLIK) may increase the overall customer satisfaction and sometimes even the security of the authentication process.

Despite the numerous opportunities, financial institutions are especially concerned about the rise of fraud in the CNP (Card Not present). This covers multitude of different attack scenarios, and requires new types of authentication and transaction authorization and risk assessment. To combat the ever rising value of fraudulent transaction in the mobile channel, institutions need tools which are compliant with the technology and built with the environment in mind. Additionally, supplying those institutions with behavioural data about the transactions and applying machine learning algorithm seem to be the state of the art fraud detection mechanism as of today. Hence, developing solutions which apply these principles is one of the most

important challenges for the sector. As the CNP fraud is often performed by stolen or malware intercepted credentials, authentication methods providing proof of presence may be a good measure to start addressing this problem.

2.3 Authentication requirements in mobile financial services

As already mentioned in Section 1.1, authentication in financial application is still mostly reliant on passwords and PIN numbers. However, due to the unique environment created by the sensors of mobile devices, physical biometric authentication had become a standard procedure as login credentials for mobile banking applications already (Sudoł & Woszczyński, 2018) - mostly by the use of fingerprint scanners. On the other however, biometric solutions utilizing different sensors such as face images are rarely used for securing transactions due to the accuracy of the methods and security of the solution that comes along with it (Sudoł & Woszczyński, 2018). These means, used in the 'traditional' process of user authentication are often burdened with a list of multiple potential issues:

- users need to remember passwords, and as it is convenient for them they use predictable and easy to remember passwords otherwise forgetting their credential causes their frustration (Lawless Research, 2016),
- users reuse their passwords and patterns, which in combination with the above mentioned tendency makes them easy to break (described more in detail in Section 3.1.1),
- physical biometrics methods encounter problems with spoofing and pattern theft, with the pattern being permanent and unchangeable,
- permanence and potential classification of a biometric pattern as sensitive data currently requires its protection. Also, the institutions must assume the pattern did not leak from one of the other services the user used it in (see Section 3.3).

Nowadays, banks and other financial institutions are more and more eager to adopt mobile biometrics, as they are safer for them and provide an additional factor for authentication, which is especially important considering the requirements PSD 2 imposes on them. Biometrics, despite their own flaws are eagerly accepted by both customers and financial institutions, as they provide additional security and may improve the usability of the service -as customer doesn't need to remember a password and process of scanning a fingerprint is shorter. As proven by

the Oxford & Mastercard conjoint study (Lovisotto et al., 2017) over 90% of customers want to adopt biometric authentication and similar fraction of banking institutions wants to adopt biometrics in their services with a similar acceptance among banks.

Adoption of biometric technologies in the mobile banking sector is promising, compared to other sectors of the economy. For example, in Poland, based on a study of bankier.pl service, over 11 banks in Poland allowed logging into banking services with the use of biometrics in 2017 (Boczoń, 2017). The updated state in Q3 2020 as of writing this dissertation is presented in the Table B.1 and accounts for 12 banks utilizing mostly more than one type of accepted biometric authentication. In most of those cases fingerprint and Apple Face ID is used (which was effectively a requirement, as new iPhone X was lacking a fingerprint scanner). As logging procedure is often allowed, transaction authorization is rare case. The only viable solution is Apple which was adopted in Alior Bank, BNP Paribas Bank, Getin Bank, mBank, Nest Bank, Bank Pekao S.A., Santander, Raiffeisen Polbank, T-Mobile Usługi Bankowe, PKO BP, ING Bank Śląski, Credit Agricole and Millennium Bank (www.dotpay.pl, 2020). This however is an major challenge, as the market share of Apple devices which support iOS is estimated at only about 25% (gs.statcounter.com, 2020). Also use of such technology makes financial institutions vulnerable to vendor lock-in problem. Google (as the biggest competitor) recently introduced its own Face ID solutions for Android devices, but as for Q2 of 2020 support for Google Face ID is lacking. There are only a few applications which support it and it is expected of them to have different performance indicators due to the difference of builtin sensors in the devices supported. Hence, based on those two problems most of the banks have been adopting their own solutions and testing various approaches, which can also be seen in the Table B.1. This proves that developing vendor independent biometric, transaction authorization viable, solutions for mobile applications an important issue from banks perspective. As for the time that this dissertation was nearly finished, some of the banks already started working on behavioural biometrics solutions, mostly relying on transactional data from their customers. This however does not belittle findings of this work, as the method proposed is unique and utilizes mostly touch biometrics, which are not mentioned in the bank's descriptions of their solutions.

2.3.1 Problems with the current authentication approaches

But despite the sector eagerness to adopt biometric solutions, current approaches do not solve many of the issues with the authentication model itself. The all-or-nothing model of authoriza-

tion which effectively requires passing the full pattern hinders usability as it requires interaction (Lawless Research, 2016). It also does not allow different levels of authorization for different services and the permanence of the pattern (such as a fingerprint scan) can be an unnecessary hindrance in terms of the security of user's sensitive data.

Banks are considering facial recognition as a new method of authentication, but this type of authentication may cause discomfort when used publicly. Utilizing your face to authenticate your transactions may prove difficult in public places or in the dark environment without sufficient lighting. This type of biometrics was still rarely used for securing transactions and mostly just for logging to the application. This partially changed with the introduction of Apple Face ID but despite the producer's assurances about the performance and high accuracy of the method face recognition have been proven to achieve performance worse than declared by the producers (Rattani & Derakhshani, 2018), especially in unfavorable lighting conditions (Günther et al., 2016; Pavlovic et al., 2018). It is also vulnerable to spoofing (Günther et al., 2016; www.androidauthority.com, 2011) by a photo or more sophisticated methods. What is the most important however, is that face recognition still requires interaction and may not be used to enrich current fraud detection systems (as archiving user's photos would be a serious privacy concern). This means that it is no different from traditional finger biometrics, and sometimes only harder to use. The change in the technology has been mostly forced by Apple and as such, remains inaccessible for the rest of the users. This in turn may point out to a question: what should be the requirements for the authentication method which does not have those negative traits?

This question is directly related with the **RG1**. What is important, is that considering the issues mentioned in the Section 2.2.5 a good method should not only provide accurate, secure and usable authentication, but also address its compliance with the mobile-centric model and as an additional benefit should provide ways to combat fraud. Also, developing such a method requires a stakeholder analysis from the perspectives of all parties which are interested in its development. This means not only banks, but also their customers need to choose this method over other allowed ways of authentication. Third parties such as FinTech companies could also use this service as a part of Open Banking architecture to increase either the security of their solutions or allowing additional access to the operations available in this channel.

2.3.2 Bank's requirements

New Payment Services Directive (**PSD 2**, Directive (EU) 2015/2366) (Europen Comission, 2015) requires Strong Customer Authentication (SCA) to strengthen transaction security and sensitive data protection. According to the directive, all banks across the European Union must add at least two-factor authentication. This means that financial institutions must comply with SCA by delivering a combination of at least two independent elements out of three categorized as: knowledge, possession or inherence factor. This requirement enables the potential use of inherence factor (biometrics) to improve current knowledge and possession based authentication. According to European Banking Authority: **inherence may include behavioural biometrics identifying the specific authorized user (EBA, 2019).**

The adoption of biometrics and similar methods in financial environment requires specific criteria to be met. The five factor framework proposed for adoption of biometrics in financial services (Lovisotto et al., 2017) considers the following elements:

- **Performance** (security the method provides) - Key performance indicators for biometrics effectiveness are commonly measured in various rates of false acceptance or error rates. The Five Factor Framework urges banks to include these KPIs but also consider the multi-level authentication that the technology provides.
- **Usability** - *Users show a preference for fingerprints compared to face recognition. Design a user experience that conveys trust and security while being easy enough.*
- **Interoperability** (technical complexity) - With the average number of user-owned devices at 4 and growing rapidly, interoperability across devices is an important challenge. Industry professionals are focused on interoperability across devices, meaning the system can authenticate users via biometrics measured by different devices (e.g., mobile phone, laptop, wearable).
- **Security** (of the pattern)- Minimize your risk by encrypting biometric templates and ensuring they never leave the user's device. Definition of threat models is one of the most important tasks when designing the security of a biometric system.
- **Privacy** - Use cutting edge protection technologies to preserve confidentiality and anonymity even within an authentication system. Biometric information leads to more personally identifiable characteristics and is a very different issue than protection of a password.

For assessing real-time transaction risk, the PSD 2 directive (Europen Comission, 2015) allows using effective risk-based approaches which ensure the safety of the payment service user's funds and personal data. This means that methods which could provide such function as an addition to authentication are of high demand. What is the most important, is that **PSD 2 directive allows bypassing the strong authentication principle in case of low risk transactions**, which includes:

- low sum transfers,
- cyclical transactions,
- transactions with whitelisted shops and institutions,
- transactions with a low risk of fraud,
- in secure corporate transactions[7].

Those cases open up a new way for not only highly usable authentication methods, but also for the risk and fraud assessment of transactions. An assessment of the risk can cause the transaction to possibly be only authenticated on the basis of user behavioural profile as an inherence factor.

Another challenge identified may be the generality of the method. Currently the lack of unified authentication method, which is connected with the number of possible combinations of phone models and built-in biometric mechanisms severely hinders evaluation of the biometric methods performance. This problem brings up an interesting research area for developing a model of authentication which is able to work regardless of the device that it is built on. This brings up a potential of using behavioural biometrics systems as a factor in a multi-factor authentication system that can work regardless of the physical biometrics sensors installed on it - as it would be based on sensors which are installed on every device.

2.3.3 User's requirements

Customer requirements have an ongoing and increasing influence on driving the innovation in banking services (Mulders & den Hertog, 2003). As it is increasingly easy to change financial services providers, churn is a serious issue. Due to the digitization of the services and interoperability of all the major platforms, the functions available to the customers are similar and prices

[7] As defined by PSD 2, https://www.eba.europa.eu/single-rule-book-qa/-/qna/view/publicId/2018_4060

of services do not differ much. This means customers opinion highly influences the banks decisions, as perceived image of the service can result in gaining or loosing customers in this highly competitive environment. This also means that usability and user experience connected with their mobile applications is important, as this channel is characterized by the highest dynamic. In addition, in the COVID era banks started to close their local branches moving customers to online and mobile services.

Main issues of user's frustration, connected with the usability of currently applied authentication solutions are either caused by forgetting the knowledge factor or the process itself (as shown in the Figure 2.9. The report itself also confirms that users find biometric solutions to be the least frustration-inducing. That doesn't mean biometrics are perfect. Based on the other VISA study from 2017, main concerns of customers about biometrics (Visa, 2017a) include:

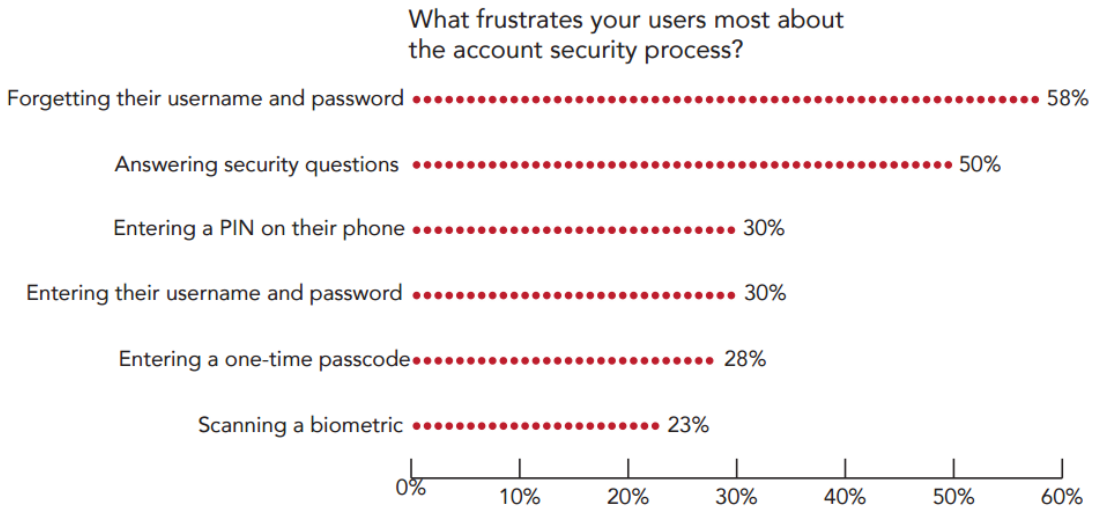


Figure 2.9. Causes of frustration during the authorization process. Source: (Lawless Research, 2016)

- 49%: **security**, leak of confidential information,
- 49%: **usability** / **accuracy of the method** issues – biometrics will take multiple tries,
- 40%: **cost of owning a device** with biometric sensor,
- 35%: **privacy** - bank having access to sensitive information,
- 20%: **discomfort** using biometrics in public,
- 20%: **no standardization** of biometric authentication.

According to the previously mentioned reports, not only the dynamics of the market itself need to be mentioned, but also the requirements and expectations of customers. According to

the VISA study, 59% of customers are worried about the security of provided solutions and 46% about are concerned with the privacy of offered services (Visa, 2017c). Even despite the consumers concerns biometrics methods are considered safe and secure not only by providers but also the consumers. 84% of respondents expresses confidence in such measures as a secure form of authentication (this confidence is confirmed also by another, earlier study (Związek Banków Polskich, 2009)). This showcases that utilization of biometric services with due diligence paid to security and convenience may be a source of major competitive advantage in the upcoming years.

Biometric identification will play an increasing role in financial services. Its success however depends on the customer acceptance. The recent study, which was trying to understand underlying customer attitudes to biometric identification, pointed out user convenience and account security as the most significant factors connected with the positive attitude towards biometrics (Mills & Zheng, 2019). However, security concerns had a negative influence in that regard. This trade-off of security and convenience (usability) is important, because users are likely to forgot their security concerns for usability of the service, provided there are enough security measures in place to protect their sensitive information. What is specially important, is to protect users from the forgery of biometric patterns. People are afraid about safety of the services they use, and want the new approaches to be both secure and providing high usability. Given the importance of perceived convenience, it is crucial that the services are easy to use minimize user actions such as the need to re-verify one's biometrics.

There is one more issue, which customer's do not emphasize directly in market studies. As proven by the Lawless Research security report (Lawless Research, 2016), the action of authenticating is itself frustrating for the user. Users tend to perform different actions on their devices with the use of their mobile financial application. According to the summary article by a large software company Netguru[8] (Samojło, 2019) not every function available is equally popular in the mobile channel, which is presented in the Figure 2.10. While some actions such as viewing account or recent transactions are very popular, ordering new financial services is not. Some of the reasons presented in the figure are tied to different levels of risks and could be grouped together in a way that accessing some operations would require stronger or weaker authentication. Similarly, despite not being popular today, some new high risk services (opening new

[8] https://www.netguru.com/

accounts, large sum transfers, loans) could require additional protection from fraud and should utilize whatever data is available to minimize that risk.

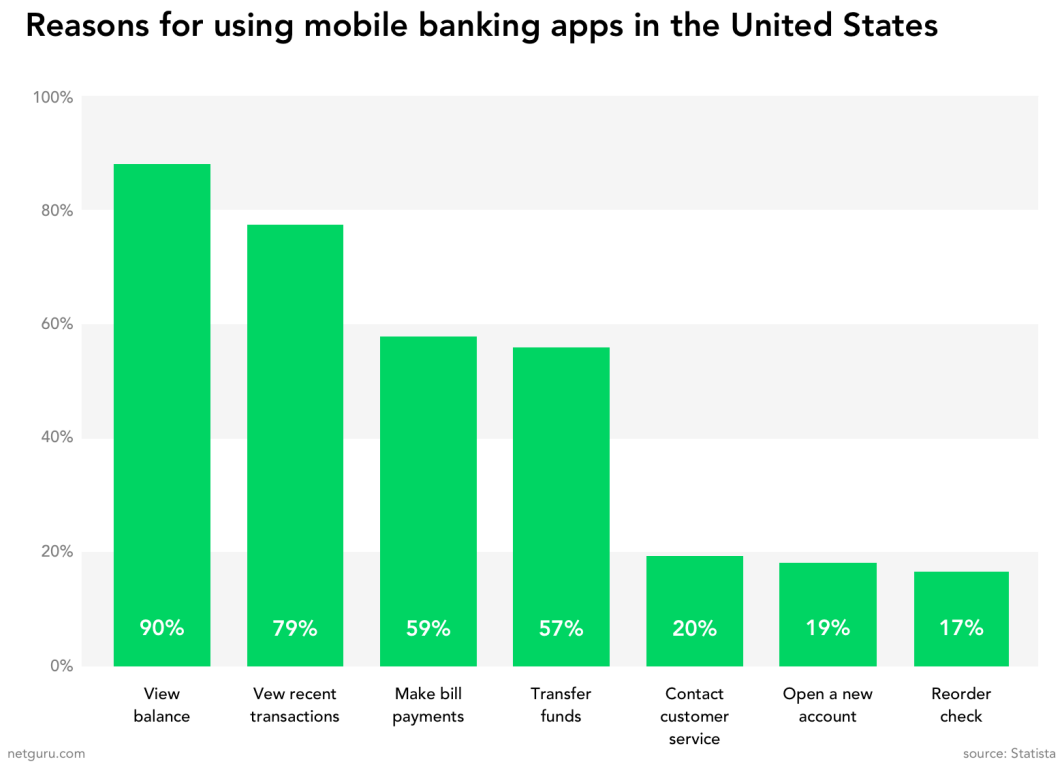


Figure 2.10. Reasons for using mobile banking app in the US. Source: (Samojło, 2019)

This lack of authorization mechanism showcases another technical issue with current mobile authentication systems. In currently used mobile application authentication procedures utilizing passwords or biometrics, the same level of permissions is used for viewing account balance, browsing recent transactions and transferring funds or opening new accounts. This is an example of the aforementioned all-or-nothing access, which allows authorization only in binary terms for the application. The change in this system may be to introduce a risk-based approach, allowing for more usable authorization for low-risk operations. Designing this new mode of authentication is another challenge for the newly developed methods.

2.3.4 Model of requirements

Based on the above mentioned examples from the literature, a model or requirements of different stakeholders can be identified. They should ensure the possibility of a successful implementation for the authentication method. The model includes the criteria of industry professionals

that can be seen in the Figure 2.11. In the dimensions of successful biometric implementation by Mastercard study (Mastercard, 2018), three distinctive groups are defined, which somewhat complies with the division of stakeholders proposed in the dissertation, consisting of: Industry Professionals (Banks), Industry Bodies and Groups (including partners or potential third parties like other financial institutions) and End Users (Customers). Among the most important dimensions that influence successful biometric implementation, the following were identified: reduced costs and compliance with the legal and technological requirements, usability and security.

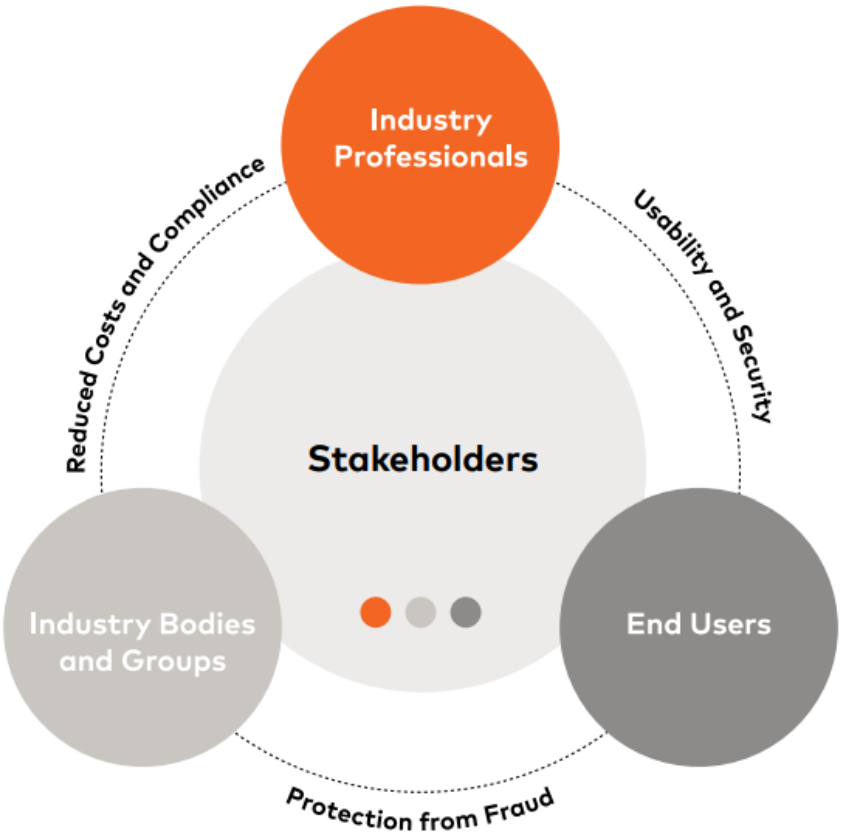


Figure 2.11. Dimensions of successful biometric implementation. Source: (Mastercard, 2018)

The previous sections of this chapter identified industry professionals (Banks) and end users (Customers) requirements. What remains to be specified are the requirements of other potential stakeholders utilizing the banking services architecture. This group mostly consists of entrepreneurs which conduct payments through the infrastructure including innovative high technology applications and services (such as from RegTech and FinTech sector) which may benefit from broadening the Open Banking environment. Technological compliance and reduction in costs of adapting to the new method, along with other benefits, can be achieved if

the method provides a certain level of interoperability. Fulfilling it could mean enabling the developed method to work with a part of partner's app (enriching the e-commerce application by bank's behavioural biometrics) and accessing the information produced by the method between stakeholders.

Summarizing, based on the findings of the chapter and the motivation presented in Section 1.1, the method proposed must comply with the following requirements, showcased in the Figure 2.12:

- (Banks/Customers/Third Parties) Security of the method **[A1]**[9] - mentioned by all of the three stakeholders identified. The method should be secure and allow for safe authentication, reliable enough for customers and banks and accessible from the third parties' perspective.
- (Customers/Banks) **[A2]** Privacy - the method should be non-privacy threatening to user data, resistant to spoofing, pattern theft and protect user's confidential information.
- (Banks) **[B1]** Fraud detection - supporting fraud detection systems and detecting insider threats, malware and unauthorized access is crucial as to the reason behind the method's implementation in the first place. Method should provide any measure which can improve current fraud detection and risk assessment systems, potentially utilizing unique information available on mobile devices.
- (Banks) **[B2]** Cost effectiveness and platform independence - the employed method should not generate substantial costs, it may not require specific devices or incur considerable costs on banking infrastructure. If the bank application is able to authenticate the user on its own, without utilizing external services, the method may meet those criteria. It should also not be reliant on a specific vendor or technology available only on a portion of the devices.
- (Banks) **[B3]** Legal requirements – developed solution should be compliant with PSD 2, with accordance to the risk based authentication on transaction and session level and also a valid authentication factor. Similarly, it should comply with the GDPR regulations in terms of protecting customer data.
- (Customers) **[C1]** Usability – it should provide high usability, possibly allowing for continuous authentication and adaptive risk based authorization that can provide different permission levels.

[9]The naming of the criteria [A] means that it is tied to multiple actors, instead of B for banks, C for customers and P for third parties. It is also referenced in the Table 1.4 that is to be used while evaluating the artifact.

- (Third Parties) **[P1]** Interoperability – the solution should be available to be easily provided in the open API banking architecture as authentication or fraud detection tools, as customers have no issues with trusting their behavioural information with banks.

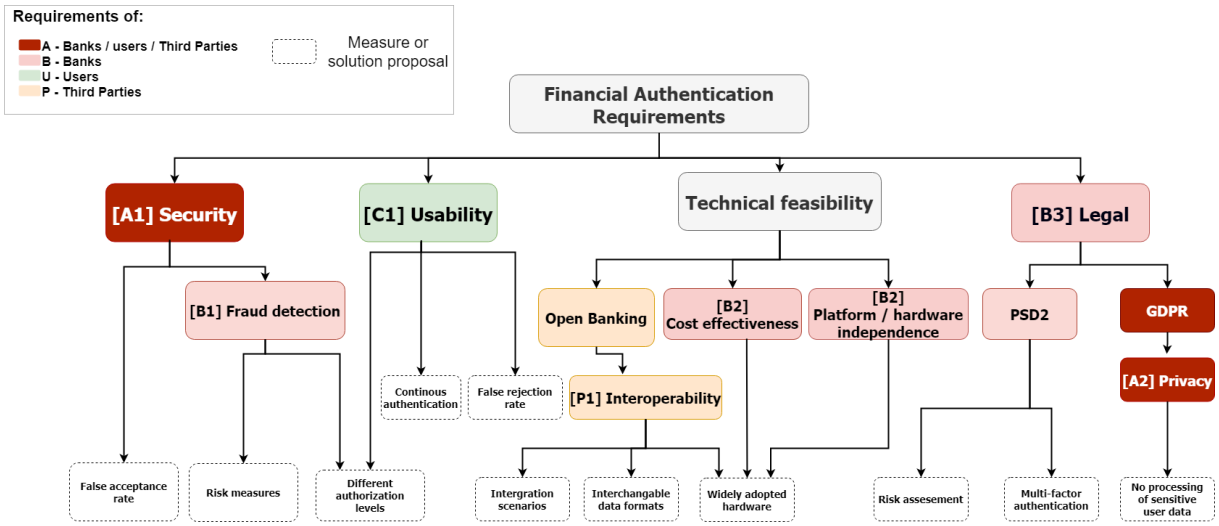


Figure 2.12. Financial services authentication method simple requirements model. Source: own development

The requirements mentioned above are connected directly with the issues most important for both the customers and financial institutions. However, the developed method would also need to comply with some of the legal and technical requirements which are implicit and/or are a standard practice in the area. Cost effectiveness and platform independence have been integrated into one requirement as they are significantly interchanged. This is mostly an effect of results presented in this chapter, namely due to the fact that hardware independent solutions may, in turn, provide low cost because they do not threaten vendor lock-in and require less sophisticated IT systems to manage and develop.

Both banks and customers value the **security and privacy** preserving characteristics of a potential method candidate. The security of the algorithm should be measured in the industry standards, where the unauthorized access probability can be measured by the average FAR (False Acceptance Ratio). On the other hand, FRR (False Rejection Ratio) showcases when the owner of the device was rejected and is tied with usability metrics. As these two values are connected and negatively correlated, a trade-off of those two values, described as EER (Equal Error Rate) is often used for quantifying the overall classifier performance.

As for the **privacy of the pattern**, the potential method should minimize the danger of pattern theft, recreation by observation and reuse in other services. As out of currently used

biometric methods both fingerprint and face scans lack this characteristic in their basic form. Hence, the development of the method devoid of this drawback may directly contribute to the solving of privacy issue. The potential disclosure of user pattern should also be considered and the method should analyze potential risks connected with it. The method should also not store or analyze sensitive user information whenever this is possible. There might also be a valid discussion, whether according to the GDPR processing of potentially sensitive data is possible. It is stated in the recital 47 of GDPR: ”the processing of personal data strictly necessary for the purposes of preventing fraud also constitutes a legitimate interest of the data controller concerned”. This means, if the method was to be used for authentication and fraud prevention, processing sensitive information is permitted if absolutely necessary. This however does not exclude the notion of minimization of possible risk to the user data. And to comply with it, method which does not require storing and processing of sensitive information would still be beneficial, if it could provide sufficient performance.

Fraud detection capabilities of the method can be synonymous to its ability to easily asses risk and provide a measure of confidence in the confirmation of user’s identity. There is a chance for new authentication methods for improving the current authentication processes beyond just the security/usability trade expressed as the number of steps needed for authentication. If the method could present a proof of presence of the rightful owner and be resistant to pattern spoofing and theft, it would protect from account sharing and unauthorized use of the device. If the method is able to provide the bank with an additional information as to the risk assessing if user is truly who she/he appears to be based on the credential’s sent, it should benefit both low risk and high risk corporate transactions by requiring different levels of authorization for different risk levels of transactions. If the method can provide this measure in near real-time, during the action of confirming the transaction, it would meet this requirement.

Providing different types of access for different actions in the application could provide an additional benefit. As not all actions are burdened with the same fraud and privacy risk, some options could be available with less time consuming process of authentication, increasing the usability in low risk transaction. This is an ongoing problem, to which some of the institutions respond with different authentication methods available for a specific action (e.g., based on risk analysis of the transaction), but these solutions can be further improved if the method could continuously asses user identity.

Expanding upon the **interoperability** requirements the data collection and authentication process should be available on nearly every mobile device and do not rely on dedicated external architecture or sensor. Both in terms of interoperability and hardware independence the method should also not rely on some specific device, as it is seen in the case of Face ID. Apple's very quick resignation from Touch ID resulted in an unprecedented and hastened implementation of such measures in financial services, which was burdened with high risk connected with the time-frame to implement such measures. Due to that, financial institutions should seek for solutions which do not incur additional costs in the devices (should be available on both Android and iOS, possibly on other OS if possible) and remain vendor independent in terms of installed sensors.

Another topic covers the communication part of the interoperability standards that the method should provide. The processing of data should be possible between banks and third parties without disclosing user information. Interchangeable data format that allows to send information regarding fraud detection and risk analysis to the banks, if possible to achieve in the developed solution, would satisfy this requirement. There are also other topics worth noting connected with the new method. Data processing for example is strongly connected with the privacy issue, its up for discussion if authentication should be processed directly by the device (in-line with the edge computing principle) or sent into a centralized authentication server. The analysis of the information shared and potential threats linked with leaking the data should be considered. Due to the extent of this work, only the possible architectural choices in this matter are presented in the validation section. This however remains an issue that should best be considered by financial institutions before employing such methods in the risk analysis process.

As for the legal requirements, method should comply with PSD 2 directive, it should especially be considered a valid authentication factor (inherence, knowledge or possession) to be used in financial services. Behavioural biometrics methods were proven to meet this criteria in Section 2.3.2 addressing the strong customer authentication (SCA) requirement and can also serve as a proof of presence authentication (Samet et al., 2019).

The usability of the process should also not be hindered, meaning the method should minimize the required interaction. It could either be less time consuming to authenticate than clicking the device (touch biometrics) or making a photo (face biometrics). The process could also take no time at and be done in background (continuous authentication). The goal is to

minimize the situations in which a user is asked for credentials. Currently the authentication always requires additional step in providing credentials, if that could be ”hidden” within a normal interaction withing the application, method should fulfill this criterion. Another issue, connected with the usability vs security trade off, is the context of use for the specific method. Current studies have proven that not all methods are applicable in every scenario (Wójtowicz & Joachimiak, 2016) and some may be undesirable in certain situations (for example face recognition in a crowded train) or perform worse in specific conditions (voice recognition in a loud public place). The issue of context awareness or context independence of the developed method (meaning it should authenticate with the same classifier accuracy and remain preferred no matter the situation) is an issue that remains to be studied.

2.4 Summary

The goal of this chapter was to provide a model of requirements for the authentication method suited for mobile financial environment inline with the **RG1**. Firstly, the chapter described the role of banking and payment systems in the current digital economy, characterizing trends and issues the financial sector is currently facing. First part focused on characterizing the market and trends shaping the services and products offered in the sector, pointing to problems connected with the conversion of the sector to the mobile-centric model as one of the main challenges for the upcoming years. Other trends observed in the sector were described, resulting in characterizing it as a highly competitive market where technological innovations play a great role. The issues connected with the authentication processes were also identified. As a result of this analysis, one of the main challenges for the sector was defined and concerns development of new, vendor independent authentication methods, which could also perform risk analysis and anti-fraud function. Later on, an analysis of stakeholder's needs identified the main characteristics that the method should have. Seven different areas were identified, valuable from the perspectives of: banks, customers and other parties connected with the sector (so-called third parties, including the OFIs and FinTech companies). The model designed directly answers the **RG1**, creating a structured form of requirements fit for the financial environment - compliant with the technological, organizational and customer's needs and designed for the mobile authentication process environment. The next goal is to review the SOTA on behavioural biometrics methods to find a method that is able to meet these criteria.

Chapter 3

Biometrics and authentication

The goal of the chapter is to present the typology of different biometric methods, with the focus on describing the characteristics of behavioural biometrics. As presented in the previous chapter, designing the authentication process for the financial environment requires a method meeting various requirements. To design such a method, a broad analysis of the literature and authentication domain is necessary to provide a method that fits this environment what is inline with **RG2**.

The first part of the chapter aims to characterize the properties of different authentication processes and methods to find desirable characteristics compliant with the requirements model of RG1. As the method's aim is to authenticate a user, the metrics used to assess the biometrics performance and accuracy are presented next. Further on, the chapter compares the traditional biometrics with their behavioural counterpart to discuss which of the pattern characteristics can be observed for both. The results of this comparison, along with the detailed description of behavioural methods further on, allow presenting the advantages and disadvantages of the methods. Literature review of different modalities presented at the end of the chapter concludes with choosing a method which is the best choice for the banking environment and its requirements, fulfilling the **RG2**. Further on the research gaps in the chosen modality are presented that the designed method must analyze to provide a feasible and properly evaluated solution for the domain. The chapter also aims at providing completeness to the developed domain structure by showcasing important constructs in the topic of authentication. By using existing literature classifications and relying on a broad presentation of different methods, the chapter also gives the developed artifact a broader perspective.

3.1 Authentication process and factors

The end goal of the developed method is to authenticate a user. To allow this process to be carried out, a pattern (or a profile) needs to be built which can be used as a representation of this user. The identity of a user can be confirmed by a distinctive and identifiable pattern, be it a PIN or a scan of a user's fingerprint. This pattern can be either very simple (password and username) or very complex (fingerprint scan data). As the pattern itself is just a captured representation of features that characterize the identity, it needs to be built first. That's why the first step in the process for providing an authentication solution should be to identify the entity and to extract a unique pattern. There are three unique processes that are connected with this notion (Bolle et al., 2013):

- **Identification** is an act of finding out what (or whom) something is. During the process we want to retrieve and quantify the characteristics of an individual or grouped features that can only be assigned to one entity. For the identification process to be valid, the features should be unique enough, so they are no two entities with the same values. The identification of an entity is a complex process that can be conducted by various methods. But in reality, where most of the real world security approaches deals with the scenario where we can tie a device with one or few potential identities, this uniqueness may be replaced by the sufficient distinguishability. Another important trait is the inability to reproduce or create the user pattern without significant prior knowledge and effort that's disproportional to the possible value achieved. Overall, the more features are included in the pattern, the more secure it becomes.
- If the identity of a user is already confirmed, **authentication** is the process of asserting that some identity claimant really is who he or she claims to be - a mechanism where the method confirms that the claimed identity is true. For example, a system can verify a personal identity by comparing a submitted biometric sample with the biometric reference template. What is important, is that in the authentication scenario the user needs to claim a chosen identity and the system conducts a one-to-one comparison to determine whether the claimed identity matches with the user template. Hence, we do not need to know any other identities (uniqueness on a global scale is not as important, although distinguishability is still highly desired). Authentication in its confirmation of identity is a verification task. To give an example, in a biometric fingerprint authentication on a lap-

top, when a user attempts to access the computer, he or she needs to type a username and then present their finger. The captured biometric sample is then compared with the fingerprint reference pattern associated with the given username. If the samples match with each other the user will be granted access, otherwise they are refused. The level up to which the pattern observed matches with the template is likely to be high, but not perfect.

- **Authorization** is the mechanism by which a system determines what level of access a user should have. It may determine access to the resources and functions available in the system. In the process of authorization, an already authenticated user is granted privileges and permission regarding the operations on resources (data, apps, native device functions). Nowadays, in mobile authentication, this is mostly done as a point of entry (all-or-nothing) approach which gives user access to all of device functions after properly writing PIN or a successful match of his fingerprint with the saved pattern. But due to the above mentioned level of deviations from the observed pattern, we might grant different privileges based on the comparison result, or allow only methods with certain level of accuracy or errors to give access to some functions.

Utilization of those three distinctive processes on the mobile device often omits the identification process and performs the authentication directly - as the user is trying to get access to a specific account or function based on the match of their saved fingerprint/PIN pattern. Depending on the failure or success of this process, the authorization is performed - be it allowing or denying the access to the functions assigned to the identity claimed by the authentication process initiator.

3.1.1 Means of authentication - factors

To perform the authentication process different features may be used. As it was stated previously, we may extract different patterns, some of them rely on a secret known by the user or his/her psychological traits. There exists a classification originating from access control systems, which divides these features into different factors used in authentication (Bolle et al., 2013; Renaud, 2005). This model consists of three groups of factors, to which we can assign a given trait. This includes: knowledge (what you know), possession (what you have) and inherence (who you are). Each of those factors groups features with similar characteristics and

showcases differences between them. These three groups of authentication factors are characterized further on.

What the user knows - knowledge factors Knowledge factor groups all of the features that rely on the memory and mental capabilities of the user they are connected with. It includes passwords, passphrases, PIN numbers and even graphical passwords. Methods of authentication relying on this category of factors have an advantage of being the most popular and cost effective method of providing security of the authentication process on the mobile devices. Modalities of this factor have a large list of potential advantages, which include:

+ The infrastructures of current IT systems are mostly reliant on knowledge factor authentication, meaning they have been tested and studied in countless examples.
+ They are hard to retrieve or copy from the user, if they are kept in secret.
+ They are simple to explain and very popular among the users.
+ Length of the passphrase can be adjusted based on the security requirement of the system.

However, despite their popularity, these methods nowadays are not sufficient for all of current authentication systems. Somewhat connected with their ubiquity these methods also have a multitude of drawbacks, which include:

– The ”stacking up” issue (Yan et al., 2004). User needs to remember multiple passphrases, is encouraged to change them regularly. Each new passphrase adds to pile of the things that the user needs to remember and may cause errors. In recent Internet users' survey from 2018 (Lord, 2018) covering over 1000 individuals, over 70% of users had more than 11 different password protected accounts. An on average, each US e-mail address is associated with 130 online accounts. This multitude of passphrases often leads to forgetting passwords, which is one of the main causes of user frustration during an authentication process, as studied by Lawless Research survey (Lawless Research, 2016).
– The security of passwords is no longer sufficient. Most passwords are easy to break or simply be guessed by an attacker. This is due to the fact that they are often reused (Bonneau et al., 2012; Doel, 2015; Lord, 2018; www.csid.com, 2012) and people use some predictable patterns like birth dates, names etc. to help them remember their passphrases. Users also tend to user very easy patterns, as proven by the study based

on a leaked data, 20 PIN combinations can open up about 27% of the devices, without any prior knowledge about the user (Datagenetics.com, 2012). Another study from 2014 (McDonnell et al., 2014) was able to unlock 50% of the PIN protected devices in less than 30 minutes.

- Possibly due to those issues, a large number of users does not lock their devices with passwords or any other means. That often leaves their applications accounts logged in, or allows finding confidential user information saved on the device. Over 40% of the users do not secure their phones at all (Fridman et al., 2015) [1]. The information collected from an unsecured phone, coupled with the issue of password reuse and using predictable patterns also significantly lowers the security of the method.
- Passphrases are very vulnerable to data breaches. They have been multiple cases of attacks which resulted in attackers gaining access to confidential information including passwords. Examples of those incidents include e.g., Equifax in 2017, eBay in 2014, Yahoo in 2014[2]. While the leak of the second category of data could not be prevented, gaining access to passwords and user logins often means giving a possibility of gaining access to other services of the user.
- They are often used in a point-of-entry authorization. Meaning, when providing a passphrase, the user has access to all of the functions of the application. This, coupled with the high requirements for the pattern security often leads to the low usability of long passwords in mobile applications. Using only PIN numbers however is very dangerous without employing additional factors due to the security of passwords issue.

These drawbacks however do not mean that using passwords is a thing of the past. They are still widely employed, and with sufficient uniqueness and complexity of the pattern, remain one of the most secure methods for authentication. From a technical standpoint with the use of hashing and the mechanisms being implemented closer to the device infrastructure it is nearly impossible to retrieve the password from the device. With the right industry-standard salting techniques employed on the authentication provider side, the passwords remain secure from the leak. But if password reuse is prominent, the leak of a password from the least secure service in which the customer uses this password compromises multiple user accounts. Due to the human nature and privacy issues in the business area, the use of password or PIN

[1] Other studies reported more 60 percent users in 2011(www.sophos.com, 2011) and 30 - 40 percent brought up by recent surveys in 2013 and 2014(Bursztein, 2014; Weisbaum, 2014; www.mcafee.com, 2013).

[2] https://en.wikipedia.org/wiki/List_of_data_breaches

protection on mobile devices is regarded as not sufficient, which has been studied by Lawless Research in 2015 (Lawless Research, 2016). Based on the survey conducted on a sample of 600 consumer account authentication professionals and security specialists in U.S. companies with 100 or more employees in 2015 conclusions were made considering the use of knowledge factor protection which are presented in the Figure 3.1. These aspects point out to a singular conclusion: passwords may be used, but they are no longer sufficient alone to protect user accounts. Passphrase authentication, based on knowledge factors (be it PIN, password or visual pattern) is the best in questionable cases, where other methods are unable to properly identify a user. It may happen due to the simple unpredictability of user behavior, errors caused by the unknown circumstances or conditions preventing the use of biometric sensors and token devices. Complex passwords still should be used for providing higher level of authorization by additionally confirming user identity.

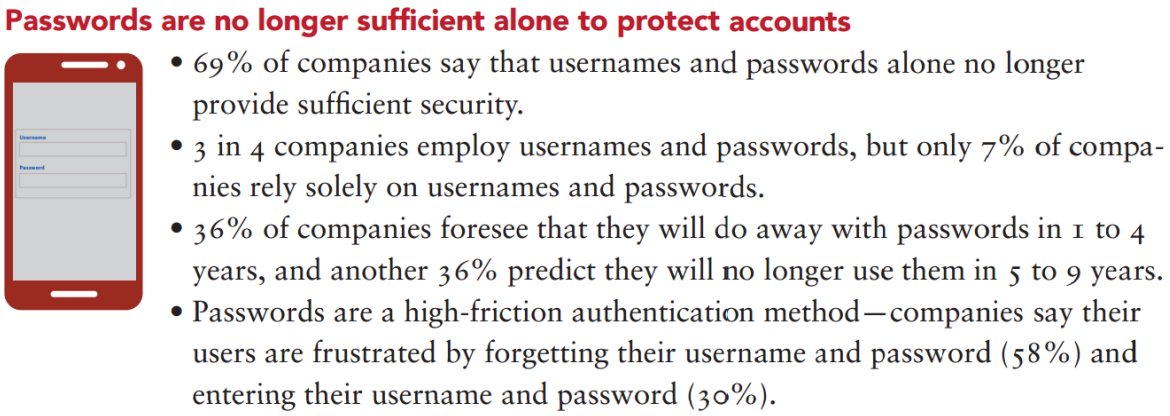


Figure 3.1. Lawless Research report findings on passwords use in account protection. Source: (Lawless Research, 2016)

What the user has - possession factor Possession factor relies on the user having access to a physical device for authentication. This family of authentication factors has gained popularity in high security areas. Sensitive or valuable data can be protected by a token (an example of the possession factor) generating device. This factor covers all of the cases, where a unique external physical object is required to authenticate the user, who possess it. It can be sometimes phrased as ”something the user and only the user has” and includes:

- Physical dedicated token devices that can generate temporary codes.
- Physical items, like a credit card used to retrieve the money from an ATM.

- Applications installed on a user's device like Google Authenticator[3] or other 2FA applications - two factor authentication solutions.

Although the concept is good for some high security ares, for everyday use it proves too complicated and cost generating. Possibly every service that requires authentication may equip the user with another device. Also, the physical access to it might be a problem in some situations, as it hinders usability. This can be partially minimized using methods that operate on the proximity of a device (similar to modern car keys). This factor is very good when security is valued far beyond the cost-effectiveness and usability. However, employing this approach is still very costly and vulnerable to the obvious human flaw of simply forgetting the device. Carrying an additional token for mobile authentication is rarely used due the cost and usability reasons. However, while carrying an additional device is uncommon in the mobile environment, often the mobile phone itself is used as a physical item fitting the possession factor principle. Token SMS codes or push notifications are indeed nothing else than using a possession factor. As this is a common practice in banking and enterprise solutions, it should draw more attention to the actual security of the device itself. As when we are relying on a mobile application installed on a device, we must assume the user (or an attacker) could already be in possession of the device.

Summarizing, the possession factor understood as a mobile device is widely used in the environment of the financial sector, but other examples of are used only in very high security cases where it is paramount to confirm the identity. Due to that, its popularity and commonness remains rather stable, partially due to the costs it may incur.

What the user is - inherence factor (biometrics) Inherence factor considers elements that are integral to the individual in question. It covers biometric traits of individuals. This could mean characteristics of either human body or actions and behaviour that differentiate people from each other (Bolle et al., 2013). This family of features seem to be a quick way to identify a user and also has a long history of use in high security environments. Biometric methods have been widely used in authentication since there were devices that would make it possible to capture fingerprint patterns. But their history is tied with forensics, and they have been used for identification far longer than that. From the biometric methods, the physical traits which are sufficiently unique among the population are used, such as: fingerprint, retina or face.

[3] https://play.google.com/store/apps/details?id=com.google.android.apps.authenticator2&hl=pl&gl=US

Although the question commonly referred to when mentioning this factor is ”what the user is?”, it also covers **what the user does (e.g., patterns, behaviour, contacts)**, answering the question ”how the user behaves?”. As not only physical traits but also behavioural ones can be captured to represent a user pattern. These methods have been mentioned as early as 2003, and have been in canon of biometrics in works such as Bolle et al. (Bolle et al., 2013). They have also been used in practical experiments for fraud detection and authentication since at least 2008 (Yampolskiy & Govindaraju, 2008). They are often classified by the literature as the second branch of the biometric methods.

The biometric methods have been proven to cause the least amount of frustration when authenticating (Lawless Research, 2016) and also have been proven to achieve fairly high accuracy metrics. Overall the characteristics of biometric features require a longer description in this dissertation, and as such, their description is continued in the next section.

3.2 Evaluation metrics of the methods

To describe biometric methods it is crucial to understand the notion of a good authentication method, especially in terms of previously described methods. There are two main factors that let us measure the quality of a biometric system:

1. How many times can someone cheat the system and gain access?
2. How many times the system will not recognize the user?

With the authentication problem we are dealing with only two classes, often labeled as 0 (the user/original possesor), and 1 (unauthorized user). The potential errors showcased in this scenario are shown in the Table 3.1.

Minimizing these two types of error is the goal of a good biometric system or method. They however may not be equal cases for the system. The threat of letting 10 illegitimate users in on 100 samples does not equal the usability cost of refusing access to the user in 1 of 10 cases. Due to the potential risks connected with the high FAR values (first case), biometric methods tend to minimize this value to an acceptable level, and only then work on decreasing the FRR (second case). The problem is that those two values are inversely correlated, and increasing one value decreases the other. Due to these facts a more detailed evaluation metrics are sometimes required to assess the method’s security.

Due to that, a few different measures need to be introduced:

Table 3.1. Classification errors in authentication.

Actual	Prediction	
	Original Possessor	Unauthorized user
Original Possessor	True Positive	False Negative
Unauthorized user	False Positive	True Negative

Source: own elaboration

- **True positives** (TP) define a situation where we classify user activity as his or hers.
- **True negative** (TN) is the properly classified activity of somebody else.
- **False positives** (FP) concern situations where the change is detected when no real change in behavior is present (in our case we deny user the access to the device because his behavior is no more similar to his pattern than the activity of another entity). From this works perspective such errors are not dangerous but annoying for the user, decreasing the usability of the system and possible adaptation. Having that in mind, additional confirmation (requiring for example a password) required should not block the use of the device for an extended period of time, but allow the user to normally use the mobile phone when on a trip to another town.
- **False negatives** (FN) which indicate not detecting the real fraud case on the other hand are very dangerous, e.g., causing the user to lose money by the unauthorized use of banking application.
- **True positive rate** (TPR) $\frac{TP}{TP+FN}$, defines the portion of found user samples.
- **False Acceptance Rate** (FAR) $\frac{FP}{TN+FP}$, in other words the portion of False Positives – letting illegitimate users / fraudsters in. This is a measure that defines the probability of letting an illegitimate user through a biometric system. It directly influences the security of the method. Therefore, an illustrative False Acceptance Rate of 1% means the system will incorrectly allow access to someone who is not allowed in 1% of cases.
- **False positive rate** $\frac{FP}{TN+FP}$ is another name for the FAR, often used in machine learning literature.
- **False Rejection Rate** (FRR) $\frac{FN}{FN+TP}$ describes how frequently we deny the access to a user. In other words it is the portion of False Negatives – rejecting the legitimate user. It is responsible for the usability of the method, as each of the unsuccessful results may cause frustration.

- **Equal error rate** - EER, sometimes called a crossover error rate, describes the point at which FAR and FRR are equal, where the lower EER indicates better performance. It can be used to summarize the performance of an authentication method in a single value. As the two mentioned above error rates are tied to each other comparing one FAR to another is not possible without knowing the FRR. And even then, due to the nonlinear nature of the methods, we need close values to be able to reliably tell which method performs better. To answer this problem an EER can be calculated, which indicates the value at which proportion of false acceptances is equal to the proportion of false rejections. The lower the equal error rate value, the higher the overall performance. The EER can be understood as a level on which the FAR equals the FRR (see bottom of the Figure 3.2). Another way of calculating the EER is reliant on the ROC curve, the intersection of the curve with the diagonal line representing the FPR/TPR equilibrium points out the EER value (right side of the Figure 3.2).

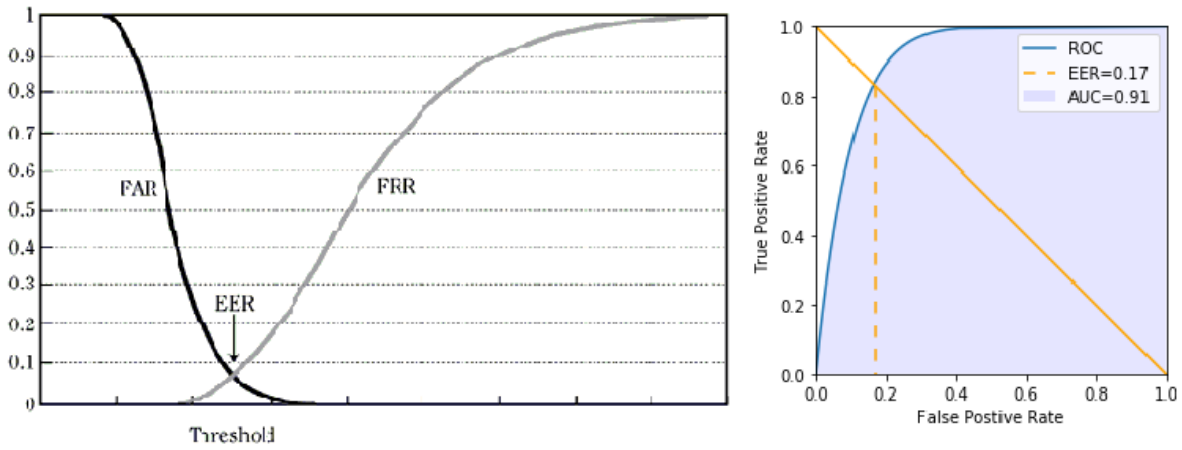


Figure 3.2. Examples of different EER representations. Source: (Bohne, 2018; Reid, 2004)

The identification process is a classification scenario, where we are choosing from one of the classes, answering a question: which user it is? Hence, to evaluate these classifiers, different metrics need to be used.

- **Accuracy** for biometrics it is often used in the literature as a descriptive metric characterizing the classifier. In machine learning and multi class classification however, it is a

metric that is the fraction of the sample the method got right: $\frac{TP+TN}{TP+TN+FP+FN}$. The name accuracy is often used as an overall description of a classifier performance[4].

- **Precision** - defined as $\frac{TP}{TP+FP}$. In our case it corresponds to the question: what proportion of positive authentications was actually correct?
- **Recall** - understood as $\frac{TP}{TP+FN}$, corresponds to the number of cases correctly identified as a user from all of the cases that the method assigned a user label.
- **F-score of F1 score metric** $2 * \frac{Precision*Recall}{Precision+Recall}$ is a metric that is used as a measurement of classifier accuracy that considers both the Precision and Recall values in similar idea to the EER. It is the harmonic mean of Precision and Recall.
- **ROC** Receiver Operating Characteristic is a curve that depicts the trade-off between TPR along the y-axis and FAR along the x-axis a various threshold values for the classification of a data instance, as is shown on the right side of the Figure 3.2. The top left corner of the plot represents the ideal point, where both errors (FAR, FRR) are equal to 0.
- **AUC** Area Under Curve is used to quantify the quality of the authentication model similarly to the Accuracy or EER. The value ranges from 0.5 to 1, where 1 represents the situations where there are no errors and 0.5 represents the random guessing. It is useful even when there is a high class imbalance between the OP and an unauthorized user/impostor.

In multi class classification we have multiple users (that are considered separate classes) and therefore most of those metrics could be skewed due to the potential class imbalance problem (e.g., X samples of positives vs. X * 100 samples of negatives). Due to that, extensions of these measures over multiple classes rely on *micro* and *macro* averages calculated for the dataset:

- micro average - micro-average aggregates the contributions of all classes to compute the average metric. Meaning it relies on number of samples in each class and weights the metric results by their contribution to the overall dataset size.
- macro average - computes the metric independently for each class and then takes the average, hence treating all classes equally.

While micro average is normally preferable if we suspect there might be a class imbalance (i.e. more examples of one class than of other classes), we should use macro average as it may

[4]Sentences such as ”X metric is a measurement of classifier accuracy” or ”we try to achieve the highest accuracy” are often used as an overall way of saying we want to minimize errors in the method and not that we have the optimization of this specific criterion in mind.

give us more information. The use of macro metrics showcases maximal errors for a specific class significantly better. It may be counter-intuitive to use a metric that shows higher errors, but it is due to the fact it provides error metrics which are closer to the ground truth. For experiments which use multiclass classification all metrics provided are tied with a macro average, as it presents the results more accurately for classes with low number of samples.

Because it ties both types of errors, EER is the best metric to compare different algorithms and methods and is used as such in this work. However, other metrics may give us a more detailed view of the characteristics of a method.

3.3 Biometrics traits

Biometric methods rely on utilizing unique (or sufficiently distinguishing) features, based on specific methods, to differentiate between people (Saeed, 2012). The traits that are connected with biometric features are multiple and include (Bolle et al., 2013; A. Jain et al., 2000; Związek Banków Polskich, 2009):

- **universality** - the feature should be observable among all of the objects that are to be identified,
- **uniqueness and distinctiveness** - the feature should be diverse enough among the objects to allow for distinction between objects. It does not always need to be unique (Kindt, 2013) (meaning we will never find an object with the same pattern) but should allow distinguishing objects efficiently based on it,
- **ease of collecting/measurability** - the processes that aims to extract and quantify the pattern from the trait should be at least possible, but it is best when they are considered easy,
- **persistence/permanence** or **stability** - the feature and its quantified values should not change with the passage of time and remain the same (or similar enough) as previously measured,
- **acceptability** - a metric tied with usability and perceived security of the method. High acceptability will be tied with methods that require less interaction, but if they use private data or have debatable ethical premises.

In the literature (Bolle et al., 2013; Kindt, 2013) other traits are mentioned which may be desirable, such as efficiency/performance - speed of the pattern extraction tied with the accu-

racy of the method. Similarly, resistance to forgery or spoofing circumvention/safety can be especially interesting from the mobile biometrics perspective. Not all of the qualities listed here are always present in a given biometric feature (A. Jain et al., 2000; Kindt, 2013). Based on the previous research, each biometric could be ranked differently according to the extent it fulfills these traits. For example, based on previous research on behavioural biometrics by (Yampolskiy & Govindaraju, 2008): *Behavioural biometrics is dependent on specific abilities possessed by different people to a different degree or not at all and so, in a general population, universality of behavioural biometrics is very low. But since the method is only applied in a specific domain, the actual universality of a given biometric may be 100%.* This approach states that universality can also be measured for the environment in which the method is used and not for the general populace.

Different biometrics may, in turn, be used based on their traits depending on the scenario of use. For example, methods with high universality can be used for governmental identification purposes, while in other cases acceptability may be the most important even when tied with higher error rates. The biggest differences in those characteristics are tied to the division between behavioural and physical biometrics. The typology of the biometric features is shown in the Figure 3.3. These methods include both physical and behavioural traits, the latter rely on a repeatable and distinctive patterns observed in the human behaviour. While the physical traits may be permanent, easy to collect and unique, behavioural traits are often only distinctive, stable (but they can change in time) and are difficult to capture and measure. Due to these differences, the characteristics of those two main groups and their influence on the authentication process is explained in the next part of the chapter.

3.3.1 Desirable traits for mobile biometrics

Considering the previously mentioned traits some are especially important in the application domain defined for this dissertation. Our research focus is centered around the financial sector, which implies high importance of pattern security. Additionally, we require that the method works well on mobile devices, improving the usability of current approaches and implies it should be available for nearly all customers utilizing this channel. Based on the summary of requirements mentioned previously, special focus is paid to the:

- Universality - as the method deals less with humans and more with ”mobile phone users”, it should use traits which do identify their owners, but also are present universally across

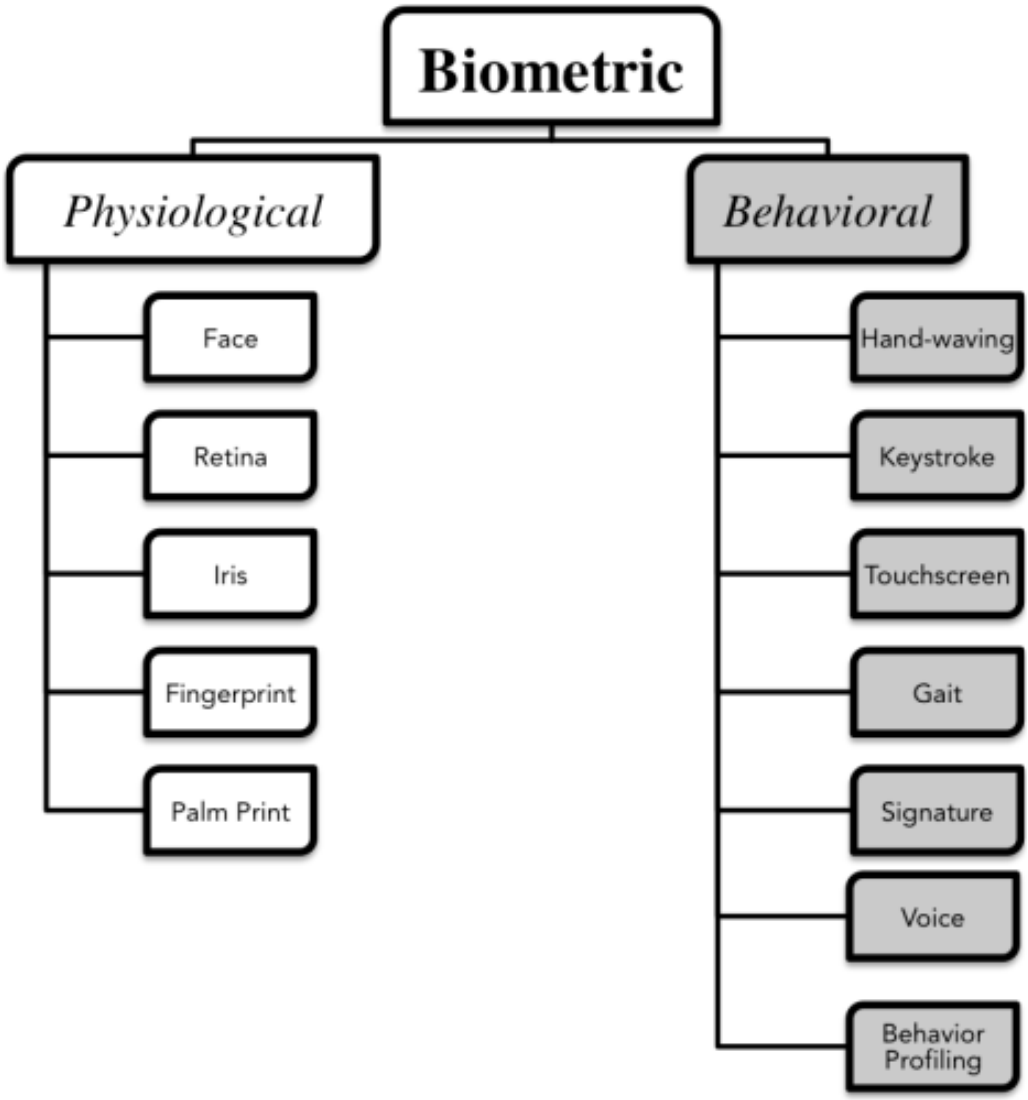


Figure 3.3. Grouping of features which can be used in behavioural biometrics. Source: (Alzubaidi & Kalita, 2016)

the devices and the features that are captured are not dependent on a very specific and rare hardware.

- Low risk of spoofing - meaning sometimes ease of collecting could be foregone, if the pattern is sufficiently hard to spoof or recreate.
- Pattern security - method should have a high security of the pattern, which have also been mentioned to be especially important in terms of successful adoption in financial services (Lovisotto et al., 2017). Not only minimizing the risk of inverting the pattern itself, but possibly minimizing the risk of pattern leak even if it happens. This could mean

non permanence may be a wanted quality. The method should be protected from pattern leak and its effects.

- High acceptability - especially in terms of processing private data.

Another potential issue is the context independence. The context independence points out to the variety of situations where the method can be successfully applied. As it has been proven in previous studies by Wójtowicz et al. (Wójtowicz & Joachimiak, 2016), not all biometrics can be used in every situation. This in turn means that the method should apply to multiple contexts or be truly context independent meaning usable in all scenarios. Methods which require specific conditions to be met or require lengthy interaction process may not fit this requirement and be rejected. No biometric method is perfect for every scenario and the mobile environment this dissertation is describing is no different. They are major issues which are connected with current biometrics methods. Those include (Sanjith, 2017):

- Noninvertibility - retrieving the biometric pattern from database kept by the authentication service provider.
- Revocability - obtaining the original pattern from multiple instances of protected biometric references derived from the same individual.
- Nonlinkability - discovering whether two samples of biometric pattern were derived from the same user.

These issues are however not equal in our environment. Revocability is not as much of an issue, as we are using the method mostly for authentication and not identification. Revocability would only be dangerous, if paired with the permanence and high overall universality of the method. Due to the fact that in physical biometrics everyone has two eyes, ten fingers etc. When an attacker somehow has an access to the original pattern (invertability issue) the revoking the pattern is no longer possible, as the original pattern that was captured could be potentially supplied. Similar principle may apply to noninvertability, which leads to the conclusion that those traits are important in two cases in our mobile environment. First is the situation where the pattern is permanent and universal. This means that obtaining a pattern from the user may cause irreparable damage and forever compromises the used pattern. Second issue is if the pattern, or the data saved with it contains sensitive private information. This might be especially important in the case of behavioural biometrics. The nonlinkability is an interesting concept in terms of privacy of the user, but it's hard to measure and plays a significantly greater role in identification than in authentication cases.

3.3.2 Physical biometrics

Out of the multiple physical biometrics that can be extracted and measured, three have been used in the mobile phone environment: fingerprint scanners, retina scanners, and facial recognition methods. Their main traits have already been covered by the previous section, namely high universality and uniqueness. Their performance in mobile environments is summarized in the Table 3.2 and particular features are described in the next paragraphs.

Table 3.2. Physical biometrics performance and feasibility characteristics.

Biometric trait	Accuracy and error rates	Ease of spoofing	Feasibility in mobile environment
Fingerprint	FAR 1%, FRR 0.000002%, 0.2 % EER possible	medium	medium (requires a sensor with sufficient performance)
Iris/Retina	FAR 0.001%, FRR 0.1%, for smartphones 1-2% EER	low	low
Face	FAR 6%, FRR 0.1%, for smartphones about 2% EER in unfavorable lighting conditions	high	high

Source: (Alonso-Fernandez et al., 2009; Günther et al., 2016; Kałużny & Stolarski, 2019; Pavlovic et al., 2018; Rattani & Derakhshani, 2018)

Fingerprint recognition - the first of those methods is also one of the most widely known biometrics. Fingerprint scans have been used for authentication and authorization on cell phones since 1998 and the error rates of these methods have decreased significantly over the years. These methods must now meet rigorous requirements("Fingerprint Unlock Security: iOS vs. Google Android (Part II)", 2016) of accuracy equaling about 99% and wrongfully authenticating an impostor in less than 1% of cases, meaning the absolute minimal requirement of 1% EER, which is often significantly lower for sensors available nowadays. Fingerprint biometric has a few distinctive advantages:

+ It has been proven to achieve low error rates.
+ Is is fast and can often be done by the press of one button on mobile devices. Despite its impact on usability (Lawless Research, 2016), it is still the lowest.

+ Is is widely accepted among the populace as already proven by the studies in Section 2.3.

However, this does not mean that it is a perfect biometric feature as is has a few drawbacks, including:

- Penetration rates - sensors for fingerprint biometrics are still not available on most of the devices, with penetration rates estimated at about 60% of all devices in 2018 (www.itweb.co.za, 2018). This means that it leaves a large portion of devices unprotected and unable to utilize this biometric method. It also leaves a large portion of the developing countries market more vulnerable due to the economical reasons of not having access to those newest devices.
- Very high danger associated with pattern leak - due to their inherent permanence, fingerprints remain relatively unchanged over the years. This, along with the fact that a single pattern (fingerprint) can be used for multiple services, means that obtaining a user fingerprint may result in the attacker gaining access to multiple user services. This also means that the weakest link principle applies, where the company with the weakest security entrusted with the pattern influences the security of the pattern itself. This issue is partially managed by the introduction of ”cancellable biometrics”, but most of the solutions are still being analyzed by researchers.
- Only point-of-entry access, inability to be used in continuous authentication - fingerprint authentication method does not allow for varying levels of authorization or continuous authentication, meaning the input of the biometric is always required on a given step of the process. As they are no fingerprint scanners installed on the screen, method can't reliably authenticate the user during normal use of the phone continuously.
- Vulnerability to spoofing - by printing the fingerprint image or pattern (Cao & Jain, 2016; Vaughan-Nichols, 2013). Quite contrary to the common belief, despite their low error rates, acquiring the pattern of a user is possible as we tend to leave partial fingerprints on any surfaces we touch. Similarly, due to the development in recent photography technologies hackers were able to retrieve a full fingerprint of Ursula von der Leyen solely based on photos (National Institute of Standards and Technology, 2014).
- Limited use in multi-modal authentication - mostly due to the interaction requirements for the authentication process.

– Sensor quality differences - due to the fact that the market of mobile fingerprint sensors is large, there might not be a standardized EER metric for these biometrics and different devices may vary in error rates observed.

Fingerprint is a viable biometric on mobile devices. However, despite its good performance and low error rates, it is highly limited by the adoption of the mobile devices. The introduction of FaceID by Apple only helped to emphasize the problem of the dependence on the sensor installed on the device, as a multitude of users was deprived of using fingerprint technology due to the vendor's decision. This, along with the inability to be used in more usable scenarios of adaptive authorization, means that alternatives to this biometric will appear.

Retinal/Iris scans - The use of unique patterns of the retina and the iris have mostly been popularized in high security environments due to their higher resistance to spoofing than fingerprint recognition. The method is not popular on mobile devices, although there are literature studies, especially connected with iris analysis (Wang et al., 2017). They have the following advantages:

+ Low error rates with sufficiently good lighting conditions and specialized hardware.
+ High resistance to spoofing attacks.
+ Potentially high penetration, if only the camera is used.

However, despite multiple publications for iris and periocular biometrics, these methods have not been widely adopted on the mobile devices, and as such are not explained in depth in this dissertation. This is mainly due to the following issues:

– Very high hardware sensor quality differences - these methods are better suited for high security stationary biometric stations or restricting physical access than for mobile devices which are equipped only with cameras and IR sensors.
– Limited usability - the act of eye scanning (be it retinal or iris scan) requires the user to remain stationary and is prone to false rejections due to the unfavorable lighting conditions and limited mobile device stabilization of the image captured.
– Very high danger of pattern leak - as eye patterns are unique to an individual and permanent.
– Only point-of-entry access - similarly to other physical biometrics, use of these methods is limited to all or nothing authentication and offers limited possibilities of continuous authentication.

As most of the drawbacks are hardware dependent, this means that the method could gain more popularity over time with the development of better cameras and popularization of smart glasses.

Facial recognition methods were first introduced to the mobile market in 2011, tied to the Samsung Galaxy phone ("Samsung facial recgonition system "Face unlock"", 2011). Despite their wide commercial introduction over 9 years ago, these methods have not yet replaced fingerprints or passphrases. They are not widely utilized among the devices because of the problems they encounter with high false rejection rates in unfavorable environments considering the lighting, environment, face position, and sometimes even race (Bo et al., 2013; Günther et al., 2016). Other reasons are tied to usability and significant battery drain. Facial recognition has gained significantly more popularity since the introduction of the IphoneX at the end of 2017. Due to this large increase in customer base and partially "forced" decision it is hard to determine its long term popularity. However, the analysis of the scientific literature on the notion of mobile face recognition has allowed to characterize its advantages and drawbacks.

+ Ease of use - face detection methods through the front camera have been quite easy to use in standard scenario of use.
+ Limited permanence - due to the effect of aging, the long term usability of a photo extracted could be limited. Despite the methods relying partially on face geometry, the leaked pattern after 10 or 20 years may prove to be less useful.

These advantages are mostly customer-focused and protect from pattern theft in the long term. This biometric however also faces some drawbacks, some of which are similar to the fingerprint biometrics:

– Very high risk and potential danger of pattern leak - photos are widespread and with the popularity of social media the privacy of one's photo (pattern) may be highly questionable. The potential cases of companies hoarding peoples' photos like ClearView AI[5] that had obtained more than 3,000,000,000 of them, point out to the fact that pattern privacy may be endangered even without the user's knowledge. Also, similarly to the fingerprint biometrics, leak of the pattern may give access to other services where user is utilizing his/her face.

[5]The company link: https://clearview.ai/. Information about stored data from: https://en.wikipedia.org/wiki/List_of_data_breaches

– Danger of spoofing - as obtaining a user photo is rather easy the nonrepudiation and security of face recognition algorithms remains highly questionable. The standard spoofing attacks could be as simple as print attack or replay attacks (shown in the Figure 3.4).
– Extremely high sensor quality differences - even more so than in case of fingerprint scanners, face recognition have been proven to be very easy to spoof and achieve high EER on most of the devices. Only Apple's FaceID introduced in late 2017 have shown to be somewhat satisfactory in terms of reported EER.
– Limited context independence - face biometrics may not be suitable to be used in an unfavorable environment, be it reliant on lighting conditions or a case of a highly populated public transport.
– Only point-of-entry access and limited use in multi-modal or continuous authentication - mostly due to the interaction requirements for the authentication process. Also, the battery drain of the camera and angle changes when using the phone (people are not always facing the phone camera) makes continuous authentication a problem. Similarly, employing the method in multi-modal authentication would hinder the usability of the process, due to the fact that the method can't authenticate the user reliably when she/he is normally interacting with the device.

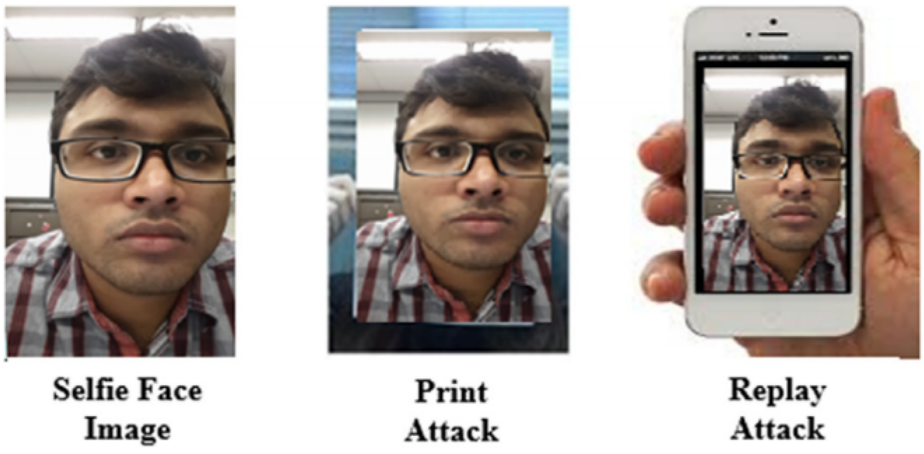


Figure 3.4. Different attacks scenarios regarding face biometrics. Source: (Rattani & Derakhshani, 2018)

Those characteristics open up a possibility of implementing methods that can act together and support the physical biometric sensors in constant authentication process and unfavorable environments or replace them in devices that are not equipped with a sensor of appropriate quality, while remaining competitive in terms of usability and accuracy compromise.

3.3.3 Reference mobile facial detection EER

Due to the fact that facial detection methods have achieved a high level of popularity in the last 2 years, a method designed should offer performance similar to it. Apple Face ID technology relying on a 3D face scan is supposed to have 0.000001% FAR, (the corresponding FRR is not listed), however the producer statement remains untested by independent studies. To the best of the author's knowledge there were no scientific papers reliably confirming this statement. There are however countless studies on the performance of the SotA methods on facial detection. Based on the results in the Table 3.2 author presented the observed EER of the mobile face detection methods in unfavorable conditions being at around 2%. While the decrease in EER is to be expected due to the introduction of IR sensor which can work in more unfavorable light conditions, and use of 3D mapping, the error rates reported by Apple do not seem to be a reliable. However, there has been a recent paper by the US National Institute of Standards and Technology that clearly stated the upper threshold of performance as of April 2020 is an EER error rate of 0,08%. (National Institute of Standards and Technology, 2020). This means that it is highly improbable that the EER in this case could potentially be better and cannot be assumed from scientific perspective, as it has not been proven in any independent studies. Also relying on recent studies of FaceID on both Apple, Samsung and other devices (Baqeel & Saeed, 2019), 4% of the users of Apple Face ID experienced problems with authentication, which could potentially point out to the much higher FRR, but due to the size of the study the final judgment on that issue also remains inconclusive. Successful spoofing attacks were confirmed for this method, they have however required printing a 3D Mask (Ramachandra et al., 2019). However, having access to multiple photos of a person the threat remains that such a model can be used for successful spoofing. Due to all of the above, the SOTA level of **EER 0,08%** will be used as a reference in the tests further on and behaviour based methods should still be studied due to the potential use in continuous authentication and authorization.

3.3.4 Behavioural biometrics

Behavioural biometrics cover unique or sufficiently distinguishable traits which can be quantified and assigned to an individual for identification and confirmation of identity - authentication (Saeed, 2012). It utilizes various methods to extract, quantify and compare the features extracted from the user to the pattern (authenticate a user) based on a set of features derived from user's behaviour. Behavioural methods have already been present and defined in the literature as far as 2003 (Bolle et al., 2013)[6], but their use and development have been significantly sped up with the appearance of mobile devices. The possibilities of utilizing a multitude of sensors available on those devices allowed for a rapid development of multiple methods. Due to this variety, those methods do not have a definitive list of features which may be used for authentication, but a literature reference typology presented in the Figure 3.3 may be used. A behavioural feature is any representation of user behaviour, its definition (also name behavioural factor can be used), which was utilized in previous work (Kałużny, 2017) is as follows:

Any readable and processable representation of user behavior which exhibits identifiable and repeatable patterns that can be used for identification and authentication.

Due to this variety, it is hard to define method's characteristics and overall performance or error rates. What is assumed by the literature however is that they may have some potential drawbacks compared to the traditional biometrics methods:

- High error rates - they may offer worse error rates in identification of an individual based on a biometric pattern, especially in a large population (Crawford & Renaud, 2014) in comparison to traditional biometrics.
- They may require multiple samples - today's biometric methods rely mostly on a one-time sample, which is sufficient as the error in capturing the pattern is low. However, due to the dynamic nature of behavioural traits, sometimes more lengthy observation is necessary to capture their uniqueness. Sometimes, it is not a single sample that gives sufficient information to confirm the person's identity, but its temporal variation (Bolle et al., 2013).

[6]The first edition of Bolle's work was published in 2003.

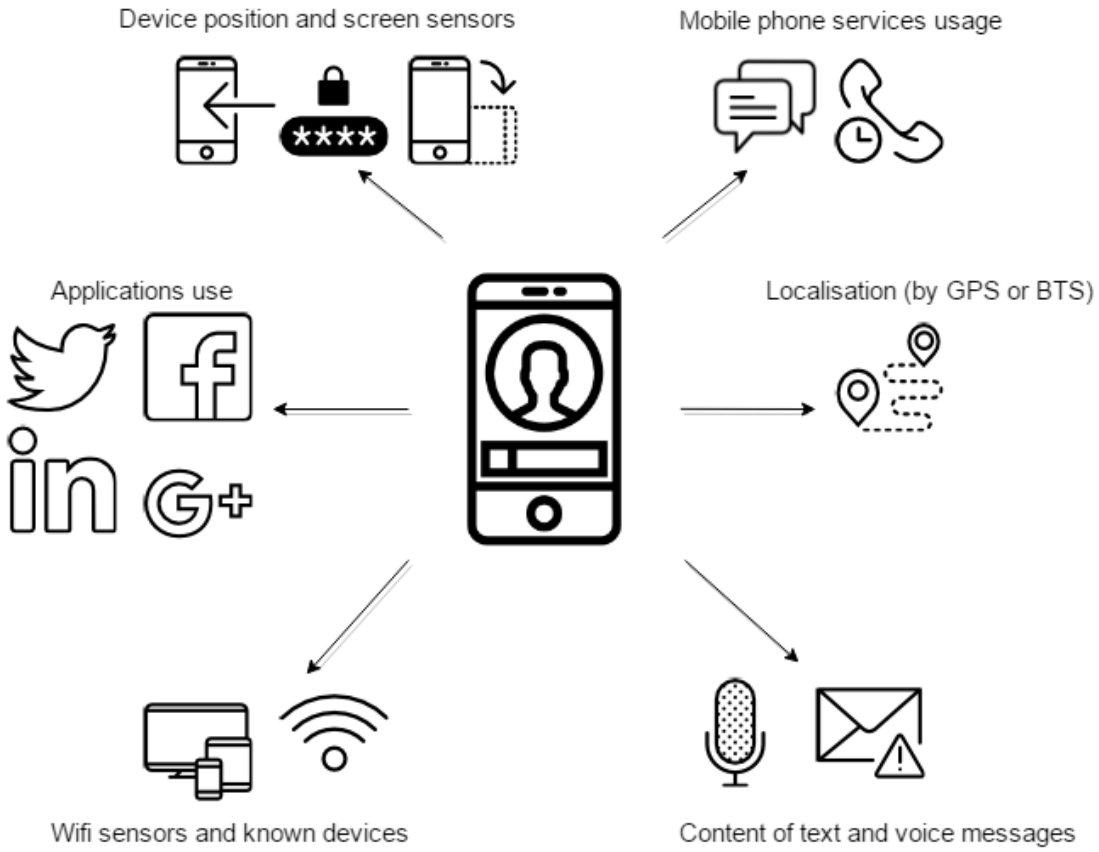


Figure 3.5. Different mobile device sensors which can be used in behavioural authentication. Source: own development

- Very high risk to privacy for some of the methods - due to the nature of capturing user behaviour, patterns may also include sensitive information. Processing localization, actions user takes outside of the application, network activity or contents of messages or voice is burdened with extremely high risk to user privacy. The leak of unaggregated data may prove to be highly concerning.
- Badly suited for the identification scenario - the identification accuracy of most behavioural biometrics is low, particularly as the number of users in the database becomes large. Their authentication (verification of identity) accuracy however can be very good (Yampolskiy & Govindaraju, 2008).

These methods also have high universality in the environment they are used in, and very small outside of it. Their uniqueness for the purpose of identification is perceived to be small (Yampolskiy & Govindaraju, 2008). They exhibit a low degree of permanence, as their behaviour may change with time. It may be stable enough to be captured, but once captured pattern needs to be updated, which largely complicates the process.

However, the family of behavioural methods can potentially offer a wide variety of potential advantages, which makes it significantly different from other methods:

+ **Continuous / implicit authentication** (Gascon et al., 2014; F. Li et al., 2014; Shi et al., 2010) - as opposed to the *point-of-entry* authentication systems, their behavioural counterpart is able to authenticate users continuously, based on the patterns captured during their interactions with the device. They can be collected non-obtrusively or even without the knowledge of the user (Yampolskiy & Govindaraju, 2008).
+ **Non-binary authorization** - connected with the above mentioned characteristics, during the process of interacting with a device method constantly produces quantifiable similarity measures between the saved behavioural pattern and the current state. By using this information it is possible define various levels of authorization based on the confidence about user's identity (Crawford & Renaud, 2014). This information can also be used to enrich fraud detection systems, even when the biometric mechanism is not the main means for authentication.
+ **Multi-layer and multi-modal authentication integration** (Bailey et al., 2014; Saevanee et al., 2012a) - Due to their varied nature, behavioural methods can easily be made from an ensemble of different behavioural features (F. Li et al., 2014) without costs in terms of added interaction or installing additional sensors.
+ **Vendor independence, high penetration rates** - behavioural methods mostly rely on the sensors already installed on every smartphone device. These sensors may vary slightly between certain producers or operating systems, but high level of consistency is required (for example in touchscreen events processing). This means that a behavioural method developed could potentially be used on multitude of devices.
+ **Cost effectiveness** - due to the no need for installing additional hardware, the methods often offer high cost-effectiveness. This may be hindered, if the method incurs considerable computational costs however. The process of data collection is fully automated and has a very low cost.
+ **High usability** of methods (Abuhamad et al., 2020; Buriro et al., 2016; Lawless Research, 2016; Xu et al., 2014) - due to their implicit nature, some of the methods can work continuously, not requiring user interaction or prompting for inputting credentials. That means high usability is to be expected. Collecting behavioural biometrics is relatively easy and

not hinders the usability. In some instances, the user may not even be aware that data collection is taking place (Yampolskiy & Govindaraju, 2008).

+ **Minimizing the danger of pattern theft** - behavioural patterns are hard to capture and measure, and sometimes patterns can only be extracted based on a prolonged observation of user behaviour. Utilizing a single or multiple methods often leads to the creation of an aggregated pattern which is not useful for the attacker as the company utilizing the methods can just change factors utilized in the process of pattern creation. There is also very low probability that multiple companies will use exactly the same behavioural features. The pattern captured also change over time (Kayacik et al., 2014) or entirely change when the user switches the mobile phone. This means that potential leak of the pattern is not burdened with a high risk.
+ **Resistant to spoofing** - It is relatively difficult to spoof the behavioural pattern as it requires knowledge of someone else's behaviour. And while for some of the methods the observation is enough, paired with inherent traits like the speed of touch, acceleration and the device specifics it is very hard to mimic the user's pattern. The pattern itself may also be kept in a form that is irreversible (such as a machine learning model or a Neural Network with a multitude of weights). Without attacker's knowledge of the system it is hard to use information retrieved from obtaining the user pattern.

Due to the large variations in behavioural methods, it is not possible to tell, if there is a method that meets the dissertation criteria without comparing different biometrics. Due to that, a further study is required to characterize different behavioural traits and methods which may be used to capture the pattern. This in turn may allow us to choose a method which fits the requirements and have the desirable characteristics.

3.4 Behavioural biometrics methods

The existence of behavioural biometrics has been widely acknowledged in the literature (Bolle et al., 2013; Saeed, 2012), even before its popularity in recent research papers. Voice and signature modalities have been studied at least as long as face or iris biometrics. Human behaviour however is complex and its different aspects may differentiate users - the task of research the behavioural biometrics is to find patterns which are unique enough to successfully distinguish people. As proven before (Bolle et al., 2013; Yampolskiy & Govindaraju, 2008) behavioural bio-

metrics creates patterns based captured representations of user behaviour achieved by the use of specific methods. There are various ways to capture behaviour, supported by multiple sensors which can be used (see the Figure 3.5) that have only been expanded since the popularity of mobile phones and wearables. Depending on the sensors quality, features extracted and the aspect captured the achieved results may vary in terms of: error rates, complexity of the pattern extraction, time required for stable pattern capture and the authentication procedure and possibilities of integration with existing systems. Listing the behaviors that can be the subjects of behavioural biometrics proves difficult due to the constantly changing research outcomes and the overlap between the methods. For example, accelerometer can be used for profiling gait pattern, recognizing user activities, as a part of behavioural profiling method or a set of variables enhancing touchscreen pattern, user signature or a written password. It can also act as a supportive sensor in geographical profiling to improve the sampling process to lower the battery usage of GPS.

Early in the literature Yampolskiy et al. (Yampolskiy & Govindaraju, 2008) characterized the earliest approaches for behaviour biometrics based on a few categories:

- authorship based - based on a pattern (text or drawing) produced explicitly by a person,
- direct human computer interaction (HCI) based biometrics - captured based on directly captured strategies and styles of the interaction with the software or hardware used,
- indirect HCI based biometrics - obtained by monitoring user's behaviour indirectly via observable low-level actions of computer software,
- motor skill based - that rely on motor-skills of the users to accomplish verification (including muscles),
- purely behavioural biometrics - measuring the human behaviour itself. Not trying to not directly capture any physically captured aspects of human behaviour (that excludes gait, touchscreen readings etc.). It would mostly cover behavioural profiling methods.

This is however not a classification, but rather a typology where each of the biometric methods can be assigned to multiple of these categories, as presented in the Table 3.3. Other possible typology of behavioural biometrics which is used in this work is based on the possible types of behavior that can be observed. This, in line with the findings of (Alzubaidi & Kalita, 2016) allows differentating:

- **Gestures** - that require a user to provide an explicit pattern which is then captured by a device. It is very similar to the traditional authentication methods such as passwords.

The gesture captured can be based on device movement (Guerra-Casanova et al., 2012) extracted by an accelerometer, camera or touchscreen sensors.

- **Keystroke** - due to the ubiquity of keyboards (virtual of physical), among both the personal computers and mobile device, capturing the way people type can be used as a behavioural biometrics. To perform this task a wide family of methods can be utilized which can also include additional sensors like an accelerometer to capture the uniqueness connected with the typing behaviour.
- **Touchscreen interaction profile** - using the sensors installed in modern mobile phones and other touchscreen equipped devices we are able to capture the process of the interaction with the device (Bo et al., 2013) and actions performed (double clicks, scrolling, phone unlocking) (L. Li et al., 2013). This includes haptic data and specific sensor readings for the details of touch.
- **Gait and activity recognition** - mostly relying on accelerometer and gyroscope measurements we can characterize the user movement (gait) and due to height, gender and unique characteristics of some users we can use these patterns to create a behavioural biometrics. This also includes specific user movement during different actions, such as sitting, running, walking, standing etc.
- **Behavioural profile** - it is one of the hardest biometrics to define, as it relies on the principle of soft biometrics mentioned by (Yampolskiy & Govindaraju, 2008). Relying on interaction patterns itself and user behaviour ”inside” the software, often including the context of place and time, it can measure uniqueness in multiple aspects of user’s daily digital behaviour. It can also be called a ”digital fingerprint” of a user, which if it is characterized by certain stability can be used to authenticate the user.
- **Signature/gestures** (or similar written pattern) - methods from this category cover authorship biometrics, which are inherently tied with the specific way a user recreates a pattern known for him or her (Diep et al., 2015). They were mostly used for digital signatures by the use of pen. Gestures on the other hand are often captured by the touchscreen device or the camera. Due to their low universality in a mobile scenario and the major decrease in using the stylus or a hand observed by a camera to interact with the smartphone, both signatures and gestures are not interesting from this work’s perspective. The touchscreen based gestures effectively utilize the same techniques as touchscreen interaction profiling, but require a defined pattern - which means they would only

add to the security achieved by touchscreen methods but they work only as a point-of-entry authentication.

- **Voice** - the unique characteristics of human voice can also be used for identification and authentication. The standard procedure assumes a predefined phrase, but the methods employed may also just try to capture patterns during voice calls. However popular, voice authentication is very highly reliant on device installed sensors, background noise and possibly privacy threatening. It has found some use in mobile banking, but researching it in large scale is very hard, hence it is not characterized by this work.

These categories of different behaviour captures do not exhaust all of the possibilities. New methods are constantly being developed and sometimes their place in this typology is unclear. For example, social behavioural biometrics (Sultana et al., 2014) based on stability of user interactions with other people. It is hard to definitely categorize it as a part of behavioural profiling or a new method. There are also multiple methods which may potentially extend this area such as: geographical aspects and the study of mobility patterns (Kałużny, 2017) or linguistic features expressed in text (Saevanee et al., 2014). Considering the variety of different behavioural biometrics and their characteristics it is crucial to analyse the literature connected with the methods that are applicable for the mobile authentication scenario.

3.4.1 Behavioural profiling

One of the behavioural biometrics which encompasses the most varying group of sensors is behavioural profiling (or behaviour profiling). Its assumptions rely on capturing implicit behaviour directly expressed in software and sensors installed instead of retrieving physical state of the user. The potential methods may use different combinations of features deriving from varying data sources, such as actions available only in particular applications. While some may be unique to the specific actions available in the software (such as calls, sms) some researchers have tried to categorize the potential features which may be extracted. For example, the work of Mazhelis et al. (Mazhelis & Puuronen, 2007) where authors described measures which can identify user behaviour. Their work produced a list of possibly obtainable information considering the nowadays smartphones generated data can be made, which consists of:

- usage of specific type of services (e.g., number of SMS sent, time spent online),

Table 3.3. Early SOTA review on behavioural biometrics.

Type of biometric	Direct HCI			Indirect HCI	Motor skill	Purely behavioural	Properties			
	Authorship	Input device interaction based	Software interaction based				Enrollment time	Verification time	Identification	Required hardware
Audit logs				●			D	D	N	Computer
Biometric sketch	●					●	M	S	N	Mouse
Blinking					●		M	S	N	Camera
Call-stack				●			D	H	N	Computer
Calling behaviour						●	D	D	N	Phone
Car driving style						●	H	M	N	Car sensors
Command line lexicon			●			●	H	H	N	Computer
Credit card use						●	D	D	N	Credit Card
Dynamic facial features					●		M	S	N	Camera
E-mail behaviour	●		●			●	D	M	N	Computer
Gait/stride					●		M	S	N	Camera
Game strategy			●			●	H	H	N	Computer
GUI Interaction				●			D	H	N	Computer
Handgrip					●		M	S	N	Gun sensors
Haptic		●			●		M	M	N	Haptic
Keystroke dynamic		●			●		M	S	N	Keyboard
Lip movement					●		M	S	N	Camera
Mouse dynamics		●			●		M	S	N	Mouse
Network traffic				●			D	D	N	Computer
Painting Style	●					●	D	D	N	Scanner
Programming style	●		●			●	H	H	N	Computer
Registry access				●			D	H	N	Computer
Signature/Handwriting					●		M	S	Y	Stylus
Storage Activity				●			D	D	N	Computer
System Calls				●			D	H	N	Computer
Tapping						●	M	S	N	Sensor
Text Authorship	●					●	H	M	N	Computer
Voice/Speech/Singing					●		M	S	Y	Microphone

D=days, H=hours, M=minutes, and S=seconds

Y - yes (possible) | N - no (not possible)

HCI - Human Computer Interaction

Source: (Yampolskiy & Govindaraju, 2008)

- social interaction patterns expressed by communication with their contacts (including time required for user to respond to an action e.g., SMS from a friend),
- geographical information including movement, locations visited and frequent routes,
- time devoted to work/entertainment or specific activities,
- way of performing tasks including sequence of actions performed on the device itself (typing, using menus and applications, touch trace),
- content expressed by the user (stylometry, voice, visited web pages),
- stability of actions and frequent patterns as the observed behaviour may have temporal patterns.

Those aspects are further connected with a set of proposed measures that can be used for the behavioural profile creation, shown in the Table 3.4. This early work done by Mazhelis and Puuronen characterized the potential features' creation process for behavioural profiling well, pinpointing temporal, geographical and social aspects expressed in the device.

The studies that led to a possibility of implementing behavioural profiling based authentication models on mobile devices are connected with the topic of fraud detection for Telecom data logs (Mazhelis & Puuronen, 2007) or personal computer use (Hilas & Sahalos, 2005). The literature often used other keywords which pointed to the same problem: "Intrusion Detection" and "Masquerader Detection". One of the first approaches used RBF neural network with one hidden layer (Boukerche & Notare, 2002) and call information (duration, number, time, date) to classify fraudulent and non fraudulent phone calls relying on CDR (Call Detail Records) data. Based on the output classification provided by a neural network[7] model was able to detect 97,5% of frauds. Other approaches also found this task can be performed utilizing even fewer features - notably the service duration(Krenker et al., 2009). Further on in 2005 (Hilas & Sahalos, 2005) an approach built user behavioural profiles based on measures derived from the same data source. The approach however vectorized those characteristics and relied on a similarity measures of user service usage and duration by comparing pattern and observed feature-created vectors.

The change in focus from the telecommunications domain to an overall authentication method with promising results is exemplified by the work of Li et al. in 2011 (F. Li et al., 2011) utilizing the MIT reality mining dataset (for details see Table C). The authors used similar temporally vectorized usage patterns of: calls, sms and mobile applications on a smartphone device.

[7]With RBF and Gaussian activation function with a mahalanobis distance metrics between the observations.

Table 3.4. List of distinctive measures proposed in Mazhelis and Puuronen article for mobile masquerader detection.

Characteristic	Measures (observable variables)
Device's facilities usage	Type of program or service evoked; temporal interval between two consecutive evocations of a program or service of a same type
Sequences of actions followed	Sequences of n actions
Temporal lengths of actions	Temporal lengths of actions
Temporal intervals between actions in a sequence	Temporal intervals between subsequent actions
Retrieving contact details from the device's memory vs entering them ad hoc.	Way of entering or retrieving contact details
Use of shortcuts vs. use of menu	For each menu command with shortcut, the chosen option
Routes taken	Sequence of cells traversed between two consecutive prolonged stops
Speed of move conditioned on route/time	Speed of move conditioned on route and time
Length of work day	Time that the terminal is in the place affiliated with the user's workplace(s); day/ time of main activities
Changes in behavior	Changes in behavioural characteristics
Words or phrases used more often	Frequency of different words used in a piece of handwriting (with stylus) or typing
Time of reading a unit of textual information	Time during which a document is open for reading
Time between incoming event and response conditioned on time of the day	Temporal interval between reading an incoming message (e.g., e-mail, SMS) and writing the response
Accuracy in typing, menu item selection, etc.	The ratio of errors to the overall number of actions, ie. the frequency of mistyped keystrokes, errors in menu item selection, ete.
Time devoted to communication	Time during a day spent for communication (using terminal) by different types of, communication (calls, e-mails, etc)
Pressure, direction, acceleration, and length of strokes	Pressure, direction, acceleration, and length of strokes
Temporal characteristics of keystrokes	Key duration time, inter-key latency time
Statistical characteristics of voice	Cepstrum coefficients ofthe signal power
People contacted with, conditioned on type of communitiation, time, etc.	Phone number, e-mail address or other address information of the contactedpeople
Places visited, conditioned on time of day, week, etc.	Locations where prolonged stops were made
Changes in the choice of environment	Changes in environmental characteristics
Time, when the user is online	Time, during which the communication facilities of the terminal are not deliberately restricted
Set of installed software	Changes of device configuration
Current screen resolution	
Volume level	

Source: (Mazhelis & Puuronen, 2007)

They classified the activities based on multiple actions and the best result achieved was an EER of 2.2% relying on text messages by building a dynamic user profile (a temporal window used for learning data) with 14 days of user's activities and 3 log entries used for classification. The error rates of a potential fusion of classifiers relying on different services was not tested in the paper. Further on better results were only achieved by Fridman et al. in 2017 (Fridman et al., 2017) which relied on an authors' dataset of 200 users' data. They built their classifier based on GPS data, application use, web browsing and partially keystroke dynamics. Relying on a 30 day period for learning data, their authentication method, utilizing a fusion of the above mentioned classifiers, achieved an equal error rate (ERR) of 5% after 1 minute of user interaction with the device and 1% after 30 minutes. The method has shown that localization plays a great role in differentiating users, especially on the shorter timespan. However, this was mostly caused by the randomization of user patterns used for fraud detection, as if we compare the user from one side of the town with randomly chosen user from the same town, their geographical patterns are different. Further studies confronting this notion (Kałużny & Filipowska, 2018) have shown that the performance of using geographical information for authentication purposes changes significantly whether the users from the dataset are from the same country, town or district. While the findings may seem obvious, they prove that methods relying on geographical features have a limited accuracy in cases of phone theft, where the fraudster is from the same town and their contribution to the classifier performance is minimal when considering insider authentication threat. Despite that, behavioural profiling has proven to achieve low error rates in the literature.

Behavioural profiling relies on the data collected by the various services and applications. The examples shown have used calls, sms, bluetooth activity, web browsing behaviour and geographical data. Due to that variety, the analysis of potential characteristics of the methods in terms of privacy, ease of use and universality is tied with the decision of using chosen services to build the profile. Judging from the study of available datasets, shown in the Table C, the data collected includes different proxies by which we can profile: social behaviour, mobility and user interests. All three of those categories can be described by the use of single or multiple sources of data. For example, social behaviour can be extracted from: calls, sms, social media application and bluetooth proximity of known and identified devices.

Overall, behavioural biometrics seem to be a highly accurate biometric, with a potential to include a nearly limitless number of variables. The methods can be verified on multiple inde-

pendent datasets, which offer extensive information about the mobile phone and are available for research purposes. Some of them contain non anonymized data (SherLock) that may be used for studying the human behaviour beyond the problem of authentication. Despite this availability, relying on the literature presented and the data collected in those datasets, we can identify several major drawbacks connected with the use of behavioural profiling method in financial applications. Those include:

- Requiring access outside of the installed application - due to the need to process geographical, network, and overall phone use information, the method would require level of access similar to the operating system. The method to work effectively requires information outside of the scope of the installed software. There are no reliable experiments in the literature that have proven that 1% EER or less can be achieved by the use of the information available on the application level.
- Access to user's private information - even if the financial application running the behavioural authentication could access all the required data only with read permissions it is still a substantial threat to a user's privacy. The methods proposed in the literature processed user calling behaviour, contacts and geographical information. While the contacts, applications and web addresses can be hashed and geographical information could be anonymized there still remains the threat that the application has access to this data. This might in turn mean that users will never know, if their private information was exposed to the provider, or some malicious party utilizing the provider's infrastructure.

The overall characteristics of the methods are presented in the Table 3.5. Summarizing the above mentioned arguments and the table's contents, the use of behavioural profiling presents a threat to a user's privacy [A2] and the potential of ensuring its interoperability and availability in the Open Banking [P1] seems to be impossible given the technical requirements of the data collection process. This does not diminish the need for studying these methods, as with the careful integration with the mobile OS and privacy preserving algorithms and procedures that would anonymize user data these methods will still be highly effective for continuous authentication.

Table 3.5. Characteristics of the behavioural profiling methods.

Characteristic		Description
Privacy	Very low	Processing highly sensitive information
Accuracy	High	Low error rates observed in the literature
Usability	Very high	Can be used in continuous authentication
Interoperability	Low	We need to assume access to observing all the OS sensors on application level, which decreases the chances of collecting data by 3rd party providers
Cost effectiveness	High	Relies on available mobile sensors
Universality (among devices)	High	Requires OS level access
Authentication speed	Medium-High	Authentication based on 1-5 minutes (session level)
Dataset Availability	Very High	Multiple datasets are available

Source: own elaboration

3.4.2 Touchscreen biometrics

Interaction patterns with graphical interfaces (GUI) and use of a haptic systems (which can collect information about direction, pressure, force, angle, speed etc.) have been mentioned in the literature on behavioural biometrics around 2005 (Yampolskiy & Govindaraju, 2008). Their characteristics stated that because so much information is available about the user interaction, a high degree of accuracy can be expected from a haptic-based biometrics system. Yet the design of those systems previously was mostly tied to the pen or sensors installed or carried directly by the user. The popularization of this topic in the literature was connected with the appearance of touchscreen equipped smartphones. Some of the first approaches can be attributed to Saevanee et al. in 2009, who enriched keystroke behavioural features with touch pressure information to achieve about 1% EER on a small sample of users (Saevanee & Bhattarakosol, 2009). Later on the focus shifted to the touchscreen gestures themselves, where Meng et al. identified actions, such as swipes and taps, characterized them and used as features distinguishing a user (Meng et al., 2012), achieving about 3% EER.

One of the most defining experiments, which created a dataset used up to today, is based on the work carried out by Frank et al. in 2013 (Frank et al., 2012) in the ”Touchalytics” project. Frank et al. identified ”trigger actions” which are frequent events which are primitive in a sense they are a part of all more complex navigational gestures. By extracting over 30 features for every swipe and stroke (2 types of identified gestures) the author achieved 0-4% inter session EER

Table 3.6. Comparison of the most widely discussed approaches for touchscreen biometrics authentication methods.

Study	# of Users	Classifiers	Feature Dimension	Performance (%)
Frank et al. (Frank et al., 2012)	41	SVM, kNN	27	EER: 0.00-4.00
Zhang et al. (Zhang et al., 2015)	50	Sparsity-based classifiers	27	EER: 0.77
Li et al. (L. Li et al., 2013)	75	SVM	10	EER 3.00
Feng et al. (Feng et al., 2012)	40	Random Forest, J48 Tree, Bayes Net	53	FAR 7.50, FRR 8.00
Serwadda et al. (Serwadda et al., 2013)	138	10 different classifiers	28	EER:10.50
Zhao et al. (Zhao et al., 2013)	78	L_1 distance	100 *x* 150 image	EER: 6.33 - 15.40

Source: (Patel et al., 2016)

for his 41 test users. Authors also discussed the differences in evaluating authentication methods for touchscreen classifiers depending on whether we include the data from one session or compare the data from separately collected sessions with each other. The EER achieved on the dataset was quite acceptable, and the possibility of improving the classifier by including inertial sensors such as accelerometer readings was proposed by Li et al. in 2013 (L. Li et al., 2013). However, the resulting performance of other studies which achieved below 1% EER were discussed by Serwadda et al. (Serwadda et al., 2013) on a dataset benchmark, which was however limited in the number of variables used for the classifiers comparisons. Overall the outcomes of different researchers prove that the observed accuracy and error rates of touchscreen biometrics methods vary in the literature, as presented in the Table 3.6. This is despite the fact that authors utilized the same aspect of user behaviour (e.g., touchscreen interaction patterns), same gestures (swipes and taps) and often similar variables. The discussion of results achieved by researchers points out the significance of other factors used in the experiment design, such as: application UI used, length of the learning process, time required for classification and the results' evaluation methodology. Some things like comparing feature importance on different datasets or showcasing the increase in performance when utilizing inertial sensors were also not discussed in the literature to the best of our knowledge, but are very important for the use of those classifiers in the real world scenarios.

Based on the previous study (Kałużny, 2019b), we can try to analyse the results described in the literature in depth. To introduce the topic, and understand the most basic differences

in the approaches, we need to describe the source of data. Extracting touchscreen events on all major mobile platforms such as Android OS and iOS is done with the help of system level APIs which can automatically classify taps, scrolls and more complicated gestures. They also provides low level information about the pressure, area of touch and exact pixels in which the centre of the activity was observed. As all of the methods depend on those APIs, the most basic data is the same inside the OS family, and the differences between them are very small due to the logic of UI and positioning required for cross platform compatibility of applications. The access to the touch events is application centered as Android and iOS prohibit access of touch data from other applications (Frank et al., 2012). This limitation creates the assumption that all of the authentication processes need to be limited to the application.

For the **events** observed, the device can monitor: x and y coordinates of the event, pressure and size registered[8]. The frequency of the data capture is not parametrizable, which may be important due to the way those devices detect action on the lowest level. The APIs differentiate between touch down events, movement and touch up events. First one appears when the user first makes a contact with the touchscreen enabled surface, then we can prompt the device to continuously monitor the movement (with certain frequency) until the user ends the contact with the device (ends the swipe) which corresponds to the touch up event. The APIs also enable accessing already preprocessed information about touch, scroll and other more complex actions. Those are also often called gestures in the literature, can be as simple as a single click (tap) where we do not expect movement events, but also include multi-touch events which require more than one finger - like zooming in. While some approaches in the literature utilize this information ”as is” available in the system API, access to the underlying events is still possible. The basic list of those gestures, along with their visual descriptions is shown in the Figure 3.6. Some of more complex gestures are still made of the simpler ones, the primitive trigger actions such as swiping (vertically or horizontally) or tapping identified by Frank et al. Since that, the use of some of the gestures is limited as during application use, swipes (also called drags/flicks/strokes) and taps occur more frequently as users browse pages to read text, or switch between the screens of an application (for example while viewing images) (Serwadda et al., 2013). Some of the actions are rarer, which means they may be less useful for building the classifier as they may not appear in every application. They may also be better suited for more complex sensors. Feng et al. in 2012 (Feng et al., 2012) studied multi-touch gestures

[8] developer.android.com/reference/android/view/MotionEvent and developer.apple.com/documentation/uikit/uitouch

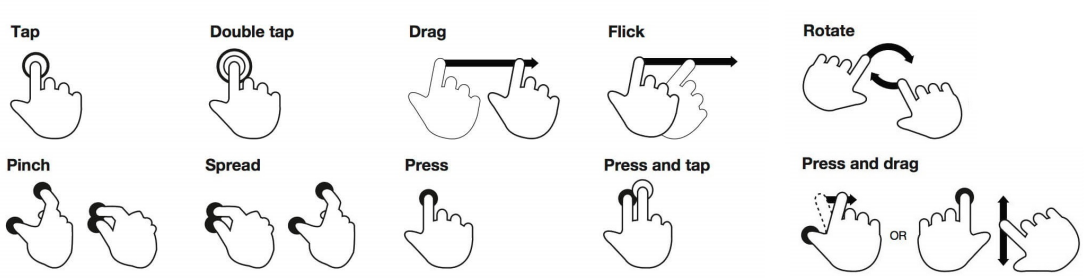


Figure 3.6. Different gestures which can be captured on mobile applications. Source: (wwww.en.profit.me, 2017)

such as: flick, pinch and spread, drag and rotate by using special gloves to identify them. The actions identified are common types of interactions with touchscreen devices, which means they apply not only to mobile phones but also smartwatches (Masood et al., 2018).

Despite the atomic API readings of the finger position, uniqueness of these gestures is also strongly connected with the stability of the interface elements the user is accessing. For example, if a confirm button is in the middle of an application screen all users would tap it, but some users may tend to click the left or right side of the button. For characterization of the touch profile uniqueness gestures that are longer than one measurement (effectively having more move event) and single measurement ones (taps) are often classified separately (Voris, 2018). To avoid errors due to the user miss clicking or performing a longer tap (which effectively becomes a swipe by the device) longer gestures consisting of only a few measurements only (<4 (Serwadda et al., 2013)) are often filtered out to increase the classifier reliability. Due to the fact that we can measure the speed and location of the touch point we can extract multiple measures, which may include the gesture's feature groups characterizing:

- **Position** - as we can measure start and end position expressed by X and Y axis pixels. This value can also be translated to relative screen position relying on the phone resolution to provide more comparative results when comparing user touch behaviour on a large scale(Frank et al., 2012).
- **Distance** - with the use of move events we can try to estimate the real distance by adding the distances expressed between the consecutive measurements of the gesture compared to the start and end-point straight line distance.
- **Time** - as each event measurement is timestamp we can find out how long the gesture took and how many measurements were observed.

- **Speed and velocity** - having the position and timestamped measurements of movement events we can often measure local speed and velocity of the gesture.
- **Area/pressure** - as the touch area and pressure is available in most of the modern mobile phones for every event, we can calculate statistics such as mean, average and the quartiles. This can be done for whole gesture but also separately for the beginning and the ending (Frank et al., 2012; Zhang et al., 2015).
- **Direction** - swipe and point curvature (A. Jain & Kanhangad, 2015) of the gesture. The mean direction for all of the move measurements included in the gesture can be calculated and expressed as an angle value. It can also then be classified to a more descriptive features such as: horizontal/vertical or one of the main 8 directions in compass-like manner (Meng et al., 2012).
- **Finger orientation and axes** and their changes - the orientation of the finger is the information available in the APIs directly and for some devices we can access separate touch minor and and major ellipse of touch point axes (Samet et al., 2019) to better characterize the specifics of user finger.
- **Phone orientation** - as the UI can change drastically dependent on the vertical or horizontal orientation so does the measurements of the gestures observed.
- **MultiTouch events** - while rare in the literature, some approaches utilize the variable defined as a number of fingers observed during the gesture (Samet et al., 2019). Limited number of publications also (A. Jain & Kanhangad, 2015) includes more complex gestures like zooming in and out.
- **Inertial sensors data** - sensors such as an accelerometer or gyroscope can characterize the way the phone is held and its position during the interaction, the approaches it may increase the classifier accuracy and decrease errors.

To characterize which variables been used by the previous researchers can explain differences in the results achieved, a literature review was carried out. Re-examination and integration of the results of the previous literature review studies (Abdulhak & Abdulaziz, 2018; Patel et al., 2016; Serwadda et al., 2013) was performed, which was complemented with the new results published after 2016. The comparison focused not only on the variables that were used, but also on the size of the sample, number of samples used for learning and classification, dataset availability and the hardware diversity used in the studies. Feature groups of the methodologically sound approaches with low error rates are shown in the Table 3.7. One of the

biggest differences between the approaches is the number of samples used for the classification. While some researchers measured their errors after a few gestures, some of them utilized a session level or even up to 70 gestures for their classification. This, along with the different designs of the test applications (as different datasets were used) may explain large differences in the achieved results. Another difference is the size of the training data in terms of number of samples and users. For example, in Li et al. (L. Li et al., 2013) 75 users were available in the dataset but the model was trained only on 28 of them, the rest was only used as impostors for testing the classifier. This table is not exhaustive in terms of the approaches, it only groups the ones that rely on gesture recognition. Other approaches were also proposed, e.g., Zhao et al. (Zhao et al., 2013) used statistics based density estimator using Graphic Touch Gesture Features (GTGF) to represent the captured touch traces as an image. Achieving 2.62% EER by combining six gestures (Zhao et al., 2013) on the sample of 30 users. What is important is that their approach was flexible enough to use multitouch gestures such as zooming, and achieved these results using 20 user traces collected over 6 sessions.

Keeping in mind that the differences in the achieved results may be caused by the learning and classification time and dataset size, the comparison on different large datasets is needed to confirm that the methods can achieve below 1% EER. This may also showcase that touchscreen biometrics are universal and the methods can be considered general, relying on the distinctive nature of features across multiple environments and dataset. What is less important in our case is proving if we can distinguish between users regardless of the task itself, as we hope to authenticate on a basis of an application. If these results can be achieved on multiple datasets with vastly different sizes a design, the approach can be used in most of the applications. The publicly available datasets and the ones which can be accessed for scientific purposes, to the best of the author's knowledge, have been presented in the Table C. The touchalytics dataset was provided by Frank et al. (Frank et al., 2012), BTAS 2013 dataset was created by Serwadda (Serwadda et al., 2013) and UMDAA02 touch dataset published in 2016 was provided by (Zhang et al., 2015). There are interesting datasets which could not be obtained at the moment of conducting this research, like the TouchMetric dataset (Samet et al., 2019), or the dataset used by Jain et al (A. Jain & Kanhangad, 2015) which are unavailable openly. Some of the behavioural profiling datasets also captured user touch patterns, for example the SherLock dataset (see Table C).

Table 3.7. Characteristics and results for the touchscreen based mobile authentication approaches.

	Position	Distance	Time	Area	Pressure	Speed	Acceleration	Direction	Finger orientation	Phone orientation	Multitouch events	Accelerometer	Gyroscope	Users	Training sessions per user	Phones	Dataset available	EER
(Saevanee & Bhattarakosol, 2009)	-	-	-	✓	-	✓	-	-	-	-	-	-		10	30	1	-	1% EER for session
(Damopoulos et al., 2013)	✓	-	✓	-	-	-	-	-	-	-	-	-	-	18	1 (24h)	18	-	0.205% EER for 24h session
(Frank et al., 2012)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	41	7	4	✓	2-4% EER for 11 gestures
(L. Li et al., 2013)	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	✓	28	600 gestures	2	-	3% EER for 14/20 gestures
(Serwadda et al., 2013)	✓	✓	✓	-	-	-	-	-	-	✓	-	✓	✓	138	80 gestures	1	✓	10% EER for 10 gestures
(Zhang et al., 2015)	✓	✓	✓	✓	-	✓	✓	✓	-	✓	-	-	-	50	3	9	✓	1% EER for 70 gestures
(Meng et al., 2012)	✓	-	✓	-	-	✓	-	✓	-	-	✓	-	-	20	6	1	-	3% EER for 10 min session
(A. Jain & Kanhangad, 2015)	✓	✓	✓	✓	-	-	-	✓	✓	-	✓	✓	✓	104	3 samples	1	-	0.31% EER for session
(Samet et al., 2019)	✓	✓	✓	✓	✓	✓	-	-	✓	✓	-	-	-	15	for 7 gestures 7	1	-	3% EER for session

Source: (Kałużny, 2019b)

Summarizing, touchscreen biometrics characteristics offers no threat to user privacy, can achieve low error rates and authenticate users after just a few simple gestures such as swipes. These methods are also widely adopted and suited for mobile devices. The characteristics of this modality have been summarized in the Table 3.8, but their applicability in terms of achieved accuracy when using only a few gestures must still be tested.

Table 3.8. Characteristics of the touchscreen profiling methods.

Characteristic		Description
Privacy	Very High	Only click patterns and accelerometer data is processed and can be aggregated on an action level (e.g., taps, swipes)
Accuracy	High	Rather low error rates observed in the literature when used in conjunction with inertial sensors
Usability	Very high	Can be used in continuous authentication
Interoperability	Very High	We assume working only on mobile phone data that is available in the system level API and inertial sensors
Cost effectiveness	High	Relying on sensors available through the system API
Universality (among devices)	High	Utilization of system level API (touch) and inertial sensors data.
Authentication speed	Very High	Authentication based on one or multiple touch actions (seconds)
Dataset Availability	Very High	Multiple large datasets spanning over multiple sessions are available

Source: own elaboration

3.4.3 Gait and activity recognition

Recognition of gait refers to the identification and authentication of an individual based a unique way they are moving. The inertial sensors installed on the devices strapped to an individual or carried by them can measure single or multi-point motion trajectories of one or multiple body segments of the subject.

The uniqueness of one's gait can be used as a biometric trait, however the uniqueness relies on capturing multiple characteristics of the movement, which can be hard to capture by utilizing only one sensor. Early work of (Derawi et al., 2010) identified distinct components that can be put together to uniquely identify person's gait. With the steps repeating in a cycle, the

identification of its specific parts allows building a person's gait profile. The study of 60 subjects who wore an accelerometer attached to a belt and placed on the left leg, by the hip resulted in an EER of 5.7%. While the value was highly promising, the authors stated that different speeds and surfaces (carpet, grass, gravel) may influence the results. Since then, multiple studies have been performed that tried to measure potential performance of this biometric. Their results have been listed in literature reviews focusing more on mobile phones (Patel et al., 2016) or wearable sensors (S. Chen et al., 2016; Mason et al., 2019). The results however do not exceed the initial error rates, with 4-10% EER reported by various studies. Due to these results, further interest in gait recognition methods may be limited due to low accuracy of the methods. Recent review (Mason et al., 2019) has pointed out: shoe sensors, smart watches and smart glasses as other potential non intrusive wearable sensors which may be used to improve the capturing process of the movement cycle.

The observed error rates are high when considering only mobile phone sensors, which are the focus of this work, relying on the results presented in the Figure 3.8 based on a literature review of Patel et al. (Patel et al., 2016). Recent approach proposed by Axente et al. (Axente et al., 2020) have designed the architecture, with both the enrollment (pattern learning) and validation phases relying solely on a mobile device sensors. However, they placed the phone in five different placements, including: torso, left arm, right arm, left leg, and right leg. Despite achieving about 90% accuracy for different activities such as walking on a flat and inclined terrain and running, use of such methodology for mobile application is not applicable. Their architecture however, presented in the Figure 3.7 is highly applicable for further development of the methods. Use of multiple anomalies for lowering authorization level and included activity recognition before the authentication process are very interesting design choices focusing on the applicability of such algorithms in the continuous authentication. The basic assumption relies on a classifier to recognize the user activity from being still and walking - which is the activity detection event. Based upon this classification a sampling service is created by the activity evaluation event to run the sampling service that validates the collected sample with the user profile. Failed validations are then labelled as anomalies, and when the counter of anomalies exceeds the limit, the gait authentication is unsuccessful, possibly blocking the access to a device.

The availability of mobile phone datasets for gait recognition remains an issue when trying to study the potential application of those methods. To the author's best knowledge, only two

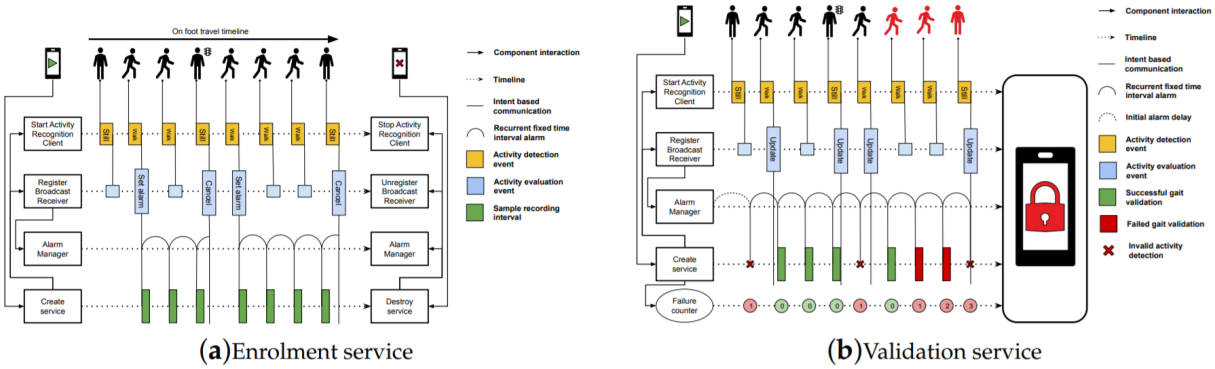


Figure 3.7. Architecture for authentication based on gait recognition. Source: (Axente et al., 2020)

datasets directly connected to this problem exist (Ngo et al., 2014) and (Panchumarthy et al., 2012). There might be also a possibility of applying datasets connected with ”activity recognition” of mobile devices to this problem. Despite that, the availability is lower than observed for behavioural profiling or touchscreen biometrics.

Despite the low accuracy the gait recognition methods, they pose a reasonably low threat to user's privacy when considering the similarity of observed patterns to the profile. However, their use in the quantitative gait analysis has become an important clinical tool and the use of such methods in various cases of assessing person's physical and mental health was listed in the literature review by (S. Chen et al., 2016). The sudden changes in human mobility and inconsistencies in the observed movement cycles can be tied with a decrease in health and can play a great role in future preventive treatments development. This is very good from the perspective of the development of such methods but may pose serious questions from the users, as the possibility of assessing their health by a financial institution may discourage them from using the method.

Summarizing, there are a few potential issues with the use of gait recognition biometrics in mobile financial applications:

- Their error rates, especially when reported for smartphone installed sensors are high as shown in the Figure 3.8.
- Despite the data being used, produced by inertial sensors, gait recognition biometrics does not directly reveal any private information. It may be used in a long term to asses the user's health, which may result in low acceptance of the method and can cause privacy issues if this type of data would be classified as health and biometric information by the GDPR.

Key gait-based continuous authentication methods for mobile devices.

Study	# of users	Feature	Classifier	Performance (%)
Mantyjarvi *et al.* [33]	36	Raw data	Correlation Coefficients	EER: 7
Thang *et al.* [37]	11	FFT	SVM	Accuracy: 92.7
Muaaz *et al.* [38]	51	Raw data	SVM	EER: 22.49 - 33.30
Nickel *et al.* [36]	48	Raw data	HMM	FNMR 10.42 @ FMR10.29
Zhong *et al.* [39]	51	GDI	Nearest Neighbor	EER: 3.88-7.22
Juefei-Xu *et al.* [35]	36	Wavelets	SVM	61.1 - 99.4 VR @ 0.1 FAR

Figure 3.8. Description of gait recognition methods results. Source: (Patel et al., 2016)

Nonetheless, the use of accelerometer and gyroscope inertial sensors, which have been the main proxy for obtaining gait patterns, have brought up a multitude of interesting methods for extracting patterns from these sensors. This had lead to the development of other behavioural biometric, enriched with inertial sensors data, which included keystroke dynamics (Giuffrida et al., 2014; Lamiche et al., 2019) and touchscreen biometrics (A. Jain & Kanhangad, 2015). In all cases, utilizing phone installed sensors have improved the classification accuracy. What is still unanswered however is the stability of the pattern observed in the long term.

3.4.4 Keystroke dynamics

Keystroke dynamics rely on identifying an individual based on his or her unique pattern of typing on the keyboard. Due to the fact that keyboards can be mechanical, digital or virtual (in case of mobile devices), the sensors that capture this biometric may rely on different types of auxiliary data. A keystroke dynamics profile is created from various typing features, such as speed, the duration between successive keys pressed, pressure applied on the keys, and finger positions. As the methods rely on registering events connected wit the keyboard, raw data needs to be processed, normalized, and stored for classification. The task of feature extraction includes mostly timing measurements capturing the duration of and intervals between the keystrokes. The timing features applied in various researches can be classified di-graphs and n-graphs. Di-graph captures the timing information between two consecutive keystrokes and can be grouped into dwell time (DT) and flight time (FT), as presented in the Figure 3.9. The dwell time corresponds to the hold time, an interval between the key press and release.

Table 3.9. Characteristics of the gait biometrics methods.

Characteristic		Description
Privacy	High	Privacy of the data is high, the patterns however my be used in a long term to asses the user's health
Accuracy	Low	High error rates observed in the literature
Usability	Very high	Can be used in continuous authentication
Interoperability	Very High	We assume working only on mobile phone data that is available in inertial sensors
Cost effectiveness	High	Relying on inertial sensors available through the system API
Universality (among devices)	High / Medium	Utilization of inertial sensors data. Universality is lower when we expect another wearable sensors to be used in conjunction with the mobile phone
Authentication speed	Medium	Authentication based on a few seconds of user cycle observation
Dataset Availability	Low	Low number of mobile focused datasets available

Source: own elaboration

Flight time is the time interval between the release of the previous key and pressing the next one. The keystroke latency is the combination of hold and flight time. Key Press Latency (KPL) is the time interval between two consecutive keys press and Key Release Latency (KRL) is the time interval between two consecutive keys release (Ali et al., 2017). The n-graph applies the same logic when considering the timing between three or more consecutive keystrokes. Different keys may have different timing, relying on the user experience with the keyboard and overall dexterity. This may include features such as: frequency of word errors (also tied to specific characters) keystroke sound, typing rate, text correction features may be included in the biometric profile. As an extension of time based variables the: position of the touch point, pressure and orientation of major and minor axes of finger-press area can be extracted for the touchscreen virtual keyboards. Additional data, such as sound of the press or accelerometer data (Lamiche et al., 2019) may also be included.

To characterize the use of keystroke dynamics overall, we may use a very extensive paper published by Ali et al. (Ali et al., 2017). The authors compared over 100 different keystroke biometrics approaches. As stated, this type of biometric has a number of advantages including

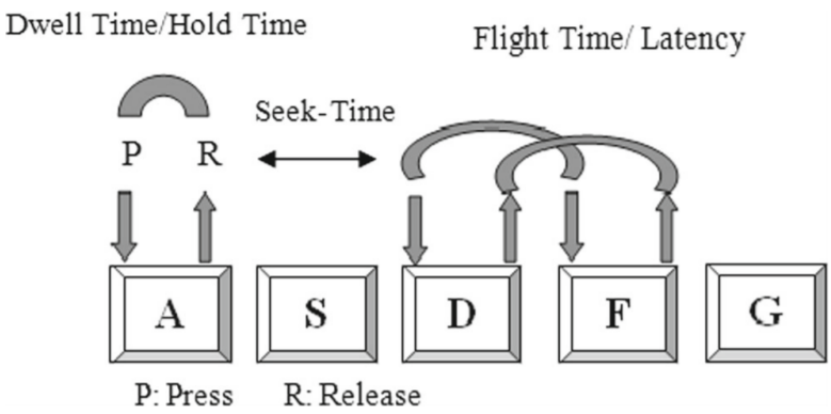


Figure 3.9. Different variables including: hold time, flight time and seek time which can be extracted from keystroke data. Source: (Ali et al., 2017)

incurring low additional costs, being transparent to the user, and non-invasive. The systems can be fully implemented in the software and have low dependency on specialized hardware as the methods rely on basic keystroke events which are required by every OS to perform basic functions. Their non obtrusiveness allows their use besides traditional passwords (during the process of a password input by user) and also continuous authentication, assuming the software requires keyboard input during the interaction process. There are however differences in potential EER achieved by different experiments as the review listed results from 0,5% EER up to about 15%. What is worth noting, approaches do not seem to offer decreasing EER over the years. This might have been caused by increasing the sample size used for authentication and hence increasing the diversity of the collected samples. What could also be the case are the different datasets used for the authentication experiment. Other factors which could influence the results may be based on: whether we observe users natural behaviour or input of a phrase in a specific application screen and how many samples were collected for one user. Higher rate of errors can also be observed, if we compare all users on the input of the same phrase or extract the features based on more lengthy observation process. That proves that the design of the authentication experiment should be tied to a situation most closely resembling potential attacks, such as user and impostors inputting a compromised password. To provide the comparability and asses this biometric performance publicly available datasets could be used. There are several available keystroke datasets, but multitude of authors still performed their experiments on their own datasets, which they rarely publish. From our perspective we are especially interested in mobile keystroke datasets, which have been presented in the Table C.3. Unfortunately these datasets provide quite a limited number of records for each user. To

provide a baseline for basic keystroke features such as DT and FT, keystroke datasets not captured on mobile devices could also be used. These include GreyC datasets[9] and other listed by the researchers in the field[10].

The SotA results achieved by researchers can be attributed to the recent results of Lamiche (Lamiche et al., 2019), which achieved about 1% EER on a sample of 20 users, while utilizing inertial sensors connected with the gait analysis along with the traditional keystroke biometrics approach. They provided an extensive analysis of the method in different scenarios, but its performance is still to be confirmed on larger datasets. This means utilization of the accelerometer and gyroscope data may increase the method's performance. Unfortunately, such data is only available for the Coakley (accelerometer, gyroscope) and Sapienta MOBIKEY datasets (aggregated accelerometer). This means that proving such high levels of accuracy can be achieved regardless of the experiment design still requires more data proven examples.

The main issue with the keystroke analytics is the technical requirement that ties it with the issue of potential privacy breaches. The implementation of such system requires authors of the application to build their own virtual keyboard on both Android and iOS. This happens due to the fact that the method needs to capture press and release times of specific keys clicked on the mobile device screen. Knowing the position of specific key presses in turn allows the application to read passwords and phrases inputted in the application. This, paired with the need of implementing own keyboard opens up issues with potential exposure of user sensitive data such as passwords. This could be counteracted by capturing anonymized keyboard events, but having access to longer text inputs by users still opens up possibilities of probability based attacks on captured patterns.

Overall keystroke analysis appears to have slightly worse performance than behavioural profiling and comparable to touchscreen analytics with the decision time being reasonably fact as in touchscreen biometrics. The potential of using this type of biometrics in the dissertation is however limited for now due to the availability of datasets and potential privacy issues of developing a virtual keyboards. They however remain a perfectly valid choice for behavioural authentication in financial environment.

[9] https://www.researchgate.net/publication/268982194_GREYC-NISLAB_Keystroke_Benchmark_Dataset

[10] https://vmonaco.com/datasets/

Table 3.10. Characteristics of the keystroke biometrics methods.

Characteristic		Description
Privacy	Low-Medium	Anonymization and aggregation is needed to prevent probability based attacks on patterns to uncover user patterns, also in case of 3rd parties that can process user input data without financial institution's knowledge
Accuracy	Medium	Low variability between sessions. Medium accuracy and EER
Usability	Very high	Use in continuous authentication
Interoperability	Medium	Reliance on a secure virtual keyboard in the application capturing the pattern and 3rd parties' compliance with security requirements
Cost effectiveness	High	Relying on sensors available through the system API
Universality (among devices)	High	Utilization of virtual keyboards relying on the OS level API
Authentication speed	High	Can authenticate based on one short phrase (e.g., password, pin), meaning few seconds are needed
Dataset Availability	Medium-Low	Low when considering mobile datasets with accelerometer data

Source: own elaboration

3.5 Choosing the method's main modality

The characteristics of all important behavioural biometrics were described in previous sections, which means we can finally choose the method which is the best for the financial services environment. Analysis of the methods in already researched examples is provided in the Table 3.12, including the results from the previously conducted research (Kałużny, 2019a).

While many methods were provided by the literature, there are also multiple topics which are not widely studied, including: the trade-off between the profile creation and accuracy, applicability of methods in different authentication scenarios and possibilities for enriching existing authentication systems (including standard biometrics (Wu et al., 2020)) and fraud detection systems. The overall findings confirm the notion that utilizing behavioural biometrics can increase the security without hindering the usability, as most of the methods can work in continuous authentication mode. These results are inline with the report of (Lawless Research, 2016), and the demand is proven by its results (see Figure 3.10). Their overall characteristics

can be seen in the Table 3.11. These modality advantages can be the most visible in the usability and security domain, with the potential to use some of the methods in proof of presence authentication to enrich fraud detection mechanisms. The drawbacks connected the methods is high EER compared to the existing solutions, potential danger to the user privacy and data for some methods and the technical complexity required. Also, the device-specific aspects of user "behavioural biometrics signature" created may limit the use cases for the methods, as the pattern may require recalculation and may change in time or when user changes the device.

Privacy is an important factor that needs to be addressed, as biometrics which focuses on the content of user communication and activity are a possible threat to user privacy. The same can be said about access to the user keyboard. Due to this, methods accessing the content spoken or written, based on linguistic profiling and those that require access to user's web or application history, should not be used as authentication methods in our case. What is interesting, is that the same can be said about voice, but this type of authentication is already used in some banks in Poland, namely Santander SA[11]. However, the interaction is done through phone connection to the bank branch, which may point out to the higher trust of the user in the physical institution he or she is calling that the modality. As behavioural profiling requires sharing user sensitive data, it is not suitable for financial environment. The methods however may be used on information logged on the application and utilize clickstreams to enrich risk assessment and fraud detection.

Methods used in the financial sector should also be able to stop the potential attacker trying to impersonate the user. They should however do it without accessing sensitive private information and be convenient to use in as many contexts as possible. The characteristics of the methods which were listed showcase that they can offer additional benefits, but the trade-offs in terms of privacy risk, convenience of use and security provided should always be considered. However the methods have proven to achieve low errors in the experiments, further tests that may prove their usefulness in mobile financial applications may be necessary.

Relying on the results presented, touchscreen biometrics has been chosen to be the best suited for the banking scenario due to the following unique traits:

- no danger of private data exposure and privacy threat,
- low error rates, possibly enriched with gait or inertial sensors data,
- high universality and cost effectiveness, low reliance on installed sensors,

[11] https://www.santander.pl

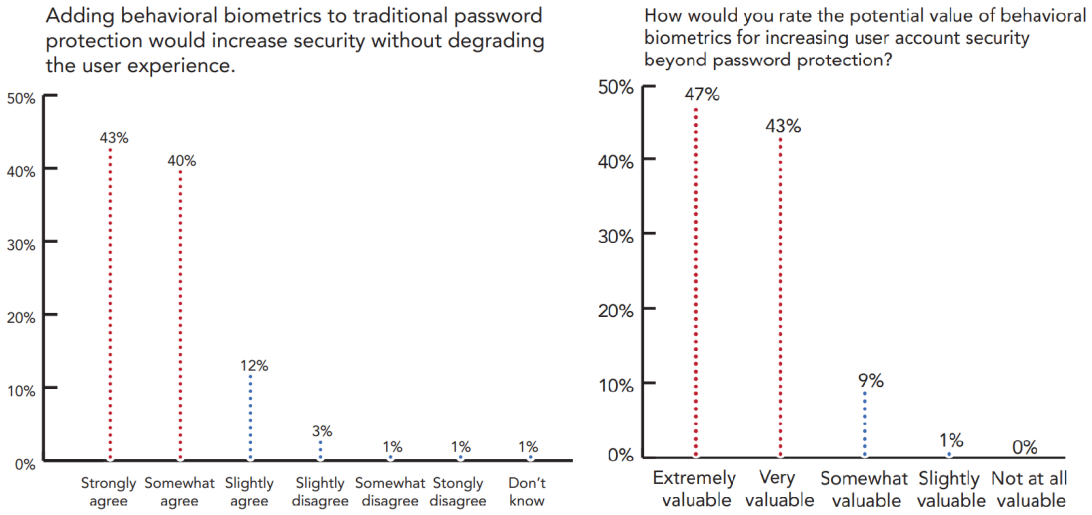


Figure 3.10. Findings of the report considering behavioural biometrics. Source: (Lawless Research, 2016)

- continuous authentication possibility.

Keystroke analysis was proven to have nearly all of the interesting characteristics, but it provides higher privacy risk, resulting in lower acceptance (bank ”tracks” what you write on your keyboard) and more technical risks involved as phone OS providers do not support custom keyboards. They are also easy to capture by attackers in repeat attacks - requiring only fake application with a virtual keyboard. They are also not as good as touchscreen profiling is in detecting bots. Most importantly, the security of user data is reliant on the 3rd party software which processes the keyboard events.

Behavioural profiling was rejected due to the fact that it achieves low error rates only when working on highly sensitive user data and requires credentials far beyond the application itself. It might be a great tool for phone OS providers, but is not feasible for implementation in financial environment, as it would require access to the user's phone data and constant data capture beyond the application.

3.5.1 Research gap - touchscreen biometrics

Touchscreen biometrics have been chosen as the most fitting for the design of financial sector applicable authentication method. The analysis of the literature presented in Section 3.4.2 pointed out important conclusions that can be drawn from comparing the approaches presented. There is no definitive answer to the extent of the method's performance. To clarify that,

Table 3.11. Summary of advantages and drawbacks of using behavioural biometrics.

Topic	Advantages	Drawbacks
Usability [C1]	Continuous authentication and proof of presence scenario applicability. May lower the number of times a user is asked for strong credentials.	Requires additional algorithms implemented in mobile application.
Security [A1]	Pattern theft is less dangerous as it can have features which are device or application specific in creating a pattern (in case of touchscreen interactions). Can provide different levels of authentication. The pattern is inherent, not a knowledge factor.	Achieved error rates of 1% EER are less secure than the fingerprint authentication. Learning user pattern takes time. The pattern may be subject to anomalies in behaviour, what requires different use case scenarios than just point-of-entry authentication.
Technical complexity (platform & hardware independence) [B2]	Methods have different requirements for hardware. Patterns are hard to spoof due to the use of multi-factor classifiers.	May increase battery usage of the application. Requires use of sophisticated algorithms and providing infrastructure.
Fraud detection [B1] detection	Can identify and differentiate user from an impostor with user credentials or malware simulating user behaviour and work as Proof of Concept. Can provide risk assessment and enrich fraud detection systems with behavioural information.	May face high False Positives ratio when detecting an impostor. Some methods may use behaviour data that might be privacy threatening.

Source: own elaboration based on (Kałużny, 2019b)

focusing on the verification methodologies for the results and the design of the experiments, two important parts of the research process highly influence the result (Kałużny, 2019b):

- **Length of the learning process and classification**, depending on how much data we collect for a particular user, and if the number of samples is spread evenly, we may create more robust classifier. Similarly, there is a large difference of what the session means in all of the studies, so authentication after a number of actions is a preferred choice to prove the method's general nature and assure comparability between the datasets.
- **Actions used for learning the classifier - the testing application design**. The design of the application in which the data was collected highly influences the results. The most important question is however, if there was an experiment design at all or users were simply observed when they performed actions in multiple applications. As our goal is to

Table 3.12. Characteristics of chosen behavioural biometrics factors performance.

Factor	Sensor	Possible use	Privacy risk	Accuracy achieved
Behavioural profiling	Call and SMS logs Application usage Battery level GPS Closest BTS WiFi, Bluetooth, NFC	Continuous authentication	High (depends on the sensors used and method)	EER 5% after 1 min, and 1% after 3 minutes (Fridman et al., 2017)
Gait and activity recognition	Accelerometer	Continuous authentication	Low	5,6% EER (Damaševičius et al., 2016)
Keystroke analysis	Virtual keyboard	Continuous authentication and enriching standard biometrics	Medium	1% EER on mobile phone enriched with accelerometer data (Lamiche et al., 2019)
Touchscreen biometrics	Touchscreen	Continuous authentication	Minimal	0,31% EER for one session enriched with accelerometer data (A. Jain & Kanhangad, 2015)
Gestures	Accelerometer, Camera	New point-of-entry method of authentication	Minimal	2% EER (Guerra-Casanova et al., 2012), 0,5% EER (Shahzad et al., 2013) with 25 training samples
Voice	Microphone	New point-of-entry method of authentication	High (depends much on user perception)	2% EER (L. Zou et al., 2015)
Signature	Touchscreen	New point-of-entry method of authentication	Low	0,8% EER (Diep et al., 2015)

Source: own elaboration and (Kałużny, 2019a)

authenticate a user inside a single application, our focus should be on datasets in which the application was designed to collect the data only within its boundaries.

Both of those elements of the research process need to be carefully analysed to properly assess the method's performance further on. Firstly, the features mentioned in the Table 3.7 may be calculated for the whole session (Meng et al., 2012), day (Damopoulos et al., 2013), or for an individual gesture (swipe). In turn, we must ask how fast we want to authenticate a user. It is a logical conclusion that we want to complete the process within a session. Some sessions in financial applications however may require more gestures, while some may be limited to a few. Due to that, we should observe the method performance after a given number of observed gestures. Secondly, we should differentiate between the available datasets, based on whether they were collected:

- accompanying a singular task: swipe or tap (Saevanee & Bhattarakosol, 2009),
- from a set of tasks, where user needs to apply multiple gestures like scrolling, (Frank et al., 2012; Zhang et al., 2015) to complete the process,
- in an uncontrolled environment (Feng et al., 2014), where the user behaviour is often observed, regardless of the application used.

The design of the application is also important as if the results are to be applicable the tasks need to be prepared in a way that they can be included in a financial application. Due to that simple actions such as swipes and taps should be used which can be reproduced in nearly every mobile application. The gestures uniqueness should be tested on multiple datasets, as it would answer if the application design is determining the uniqueness of the features for specific task, or are they inherently different between users regardless of the scenario.

Another variable that may influence the results, which will however be limited by the datasets available, is the number and the length of sessions for each user. As we may ask the user to perform one session, often in the same day and recognize the actions taken randomly from the data (inter-session) or use test data collected a week/month etc. later to confirm the pattern stability. However this poses a problem as datasets differ in the number of sessions per users and time span in which they were collected. To compare the approaches inter-session approach or randomly taking the data from all sessions can be used and the stability of the pattern can be subject of a further study. If we utilize sessions from multiple datasets however and achieve comparable results, we may prove that the method performance is not highly influenced by the inter session changes. Comparison of the results with the existing approaches

should take into consideration these characteristics, for which the different available datasets may be a good benchmark of performance. It should also answer the question of whether the importance of specific features can be attributed across multiple designs and application, which would prove the universality of this biometric in the mobile authentication scenario.

3.6 Summary

Summarizing the outcome of this chapter, possible sensors and methods were analysed to complete the **RG2** which was concerned with listing possible sensors, methods and combinations that could be used for mobile financial applications behavioural authentication. As an answer to that, each of the behavioural biometrics features listed was characterized based on the available literature. This analysis was preceded by showcasing main concepts from the authentication domain, describing the processes, metrics and traits of methods which may be used for identification, authentication and authorization processes in Sections 3.1 and 3.2. Finally, after a detailed study of the literature on each modality advantages and drawbacks presented in Sections 3.3 and 3.4 , mobile touchscreen behavioural biometrics methods were chosen as the best fitting for the requirement's of the financial sector, thus answering the **RG2**. The descriptions were aimed at assessing their usefulness in the mobile environment, and as such special attention was paid to the results in the literature that were carried out on mobile phone data. Each of the biometrics was compared with the RG1 requirements model defined in Chapter 2.

The chapter also partially answered the **RG3**, which was directly connected with choosing touchscreen biometrics as the most suitable family of methods regarding the aforementioned model. Characteristics of this biometric have significant advantages over other behavioural biometrics in the areas of [B1] fraud detection, [A2] privacy, and [B2] cost effectiveness criteria of the requirements model. The features which can characterize user touchscreen profile were listed and assigned to one of 10 different categories. The use of these features in various studies in the literature was presented along with the error rates achieved. The methods required to extract each category of features in the available datasets were analysed along with the design constraints of the experiments in which the data was collected. Possible issues which need to be answered further on in the conducted experiments, and were not answered by the previous studies in the literature, were presented in Section 3.5.1.

Chapter 4

Method for behavioural biometrics authentication

The main purpose of the chapter is to design a biometric authentication method which suits the requirements for the use in mobile financial services environment. It presents the design choices that aim to achieve performance and usability characteristics required by the requirements model. The detailed description of the method and its underlying features are also presented. To verify that it meets the accuracy criteria, a design of the verification experiment needs to be described along with the potential issues that may lie outside of its scope. The evaluation criteria are then explained further in the chapter, along with the metrics used for the evaluation and validation of these criteria. In the last part of the chapter the underlying classifiers are described, along with the hyperparameter optimization strategies and loss functions used in the method's learning process.

The chapter contributes to the overall goal of the dissertation by providing the main artifact mentioned in the **RG4** - an authentication method meeting the criteria for security and usability required for its use in the financial services environment. According to the literature reviewed in Section 3.3.4 and the answer to the RG2, touchscreen biometrics were chosen as the modality used as a base for the authentication method's design. They seem to be the best suited to the characteristics of financial environment according to the SOTA analysis done in the previous chapter.

In this chapter the measurable criteria required by the requirements model of RG1 are designed for the evaluation of the method. To define them, the list of touchscreen biometrics characteristics and feature groups in the previous chapter is expanded upon and turned into

variables that are possible to be extracted from the data. Descriptions for the calculation of the individual features that are used in the method are provided. This aims at partially meeting the criteria of **RG3** in terms of characterizing a list of touchscreen biometric features for user identification and authentication, that is evaluated in the next chapter.

4.1 Requirements

As summarized by the describing challenges of the financial sector in the motivation of the thesis, the main issue of the research is that financial services require authentication methods suited for mobile application environment, which could enrich current fraud-detection systems on transaction level, while retaining the security and improving the usability of the process compared to the currently used methods. As described before, the created method's goal is to authenticate the user based on data generated by universally installed device sensors and suited for mobile application. To meet these requirements, the developed artifact needs to achieve the criteria of the requirements model created by the RG1 (security, usability etc.). According to the desired characteristics it should retain the security offered by other mobile authentication solutions and improve the usability of the process. The designed method should:

- In accordance with the RG3 utilize combined performance of multiple features described in the previous chapter and chose features which can allow effective authentication of users. Choosing the right combination of features aims at benefiting the classifier accuracy, minimizing the error rates and meet the requirements for use in financial services defined by RG1 in Chapter 2, Section 2.3.4.
- Provide performance metrics (accuracy, error rates) comparable or better than the existing methods of mobile face recognition as stated in Section 3.3.3 and [A1] criteria of the requirements model in RG1, along with other required characteristics, such as preserving the privacy of data, as summarized in Section 3.5.
- Fit well into a user-centric and mobile-centric banking applications environment. According to the analysis conducted in the previous chapter, answering the RG2, touchscreen biometrics holds the most promising characteristics and is the best fit for mobile financial services authentication. It provides the possibility of continuous and proof of

presence authentication and can supply additional information to the fraud detection mechanisms.

The way in which every criteria from the requirements model is evaluated is shown in the Table 4.1. The performance metrics achieved, namely represented by the EER of the method is to be compared with reference mobile facial detection EER described in Section 3.3.3, but also with other behavioural biometrics methods shown in Table 3.12. The methods ability to correctly compare actions to user profile and reject fraudulent ones is a proof that it can work in a fraud detection scenario. Outputting probability or similarity measure to user profile may also be used for risk assessment of the transaction.

Despite the security [A1] metrics importance, the proposed method must also provide other characteristics, which are required to meet specific financial sector requirements. Namely, these include:

- Usability [C1] - possibility of employing continuous authentication and lowering the number of required point-of-entry prompts for the user in cases where we are sure about the user's identity due to the successful authentication while observing the user normal behaviour in the application.
- Fraud detection [B1] - by providing non binary, adaptive authorization and the risk assignment connected with the performed action. This would allow enriching the fraud detection systems and provide a risk assessment function on a transaction level, independent of the value of said operation.
- Cost effectiveness and platform independence [B2] - designing a method that works on mobile device, inside of an application (no interference with mobile OS required, as all data is provided by essential mobile OS system API) and does not require additional hardware sensors. This would prove that additional costs of employing such application are marginal.
- Legal Requirements [B3] - the use of behavioural biometrics has already been confirmed to be a valid factor in an authentication scenario by the literature, as presented previously in Section 2.3.2. Secondly, those methods have already been recognized as valid inherence factor examples for PSD 2 (Europen Comission, 2015). On the other hand, possibility of using method output as an effective risk-based approach, which could ensure the safety of the payment service user's funds and personal data is still to be demon-

strated by the implementation scenario possibilities. Similarly, the explainability of the method's results need to be showcased to ensure compliance with the GDPR.

- Interoperability [P1] - by enabling the integration with current authentication systems and the feasibility of the method's implementation in the Open Banking architecture. The validation of the method in financial use case scenarios provides proof that the method can meet this requirement. This includes: the description of the method's adoption into a mobile financial application, proposed architectural designs which consider the privacy and performance constraints, along with the data formats used for communication with fraud detection systems that might be used, which should also partially fulfill requirement of privacy criteria [A2].
- Privacy [A2] - based on the findings from the literature review, summarized in the Table 3.12, we see that touch profile dynamics and the accelerometer data provide the lowest privacy risk from all of the behavioural modalities. Touchscreen biometrics processes mostly event information and contains data only about the touch events and inertial sensors readings, which are not sensitive user information and can be aggregated. However, the possibility of ensuring the security of the data model and processes used for authentication and communication between the banks and other financial institutions remains to be proven.

4.2 Method's design

The designed authentication method will consist of a machine learning classifier that models user behaviour as an either anomaly and potential fraud, or fitting the previously created pattern. Its design will be reliant on the touchscreen events data supplied by the API operating directly under the mobile operating system to extract basic touch information as the application is running. Capturing touch events that are produced during user interaction it enters the learning phase, supplied with the impostor (other users) data. The method design consists of the following steps:

1. **Data collection and processing** - data in the format of touch API system events (as described in Section 3.4.2) are processed and the features describing user actions are extracted from the basic data format in a way showcased in Section 4.4.

Table 4.1. Requirements and criteria for method's evaluation.

Banking model requirement, Section 2.3	Requirement measurement	Criteria	Corresponding experiment section
Security [A1]	EER (Equal Error Rate) - connected with the FAR (False Acceptance Rate).	$\leq$ 0,08 % EER.	Classification scenario in Experiment 2, Section 5.2.2. Authentication scenario proposal in Experiment 3, Section 5.3.
Fraud detection [B1]	Possibility to output probability or similarity metrics on transaction level.	Binary (Yes or No), confirmed with proof of concept example.	Experiment 3, Section 5.3.1. Data model and deployment scenarios in validation, Section 5.8.2.
Cost effectiveness and platform independence [B2]	Requiring only system API information and not incurring additional costs on sensors.	Binary (Yes or No), confirmed with proof of concept example that the method can work in a single application and relying on sensors available on nearly every smartphone.	Scenarios of deployment along with the data model in Section 5.8.2.
Legal requirements [B3]	Meeting the PSD 2 requirements for: authentication factor and risk detection. Meeting the GDPR requirements for providing explainability of decisions taken by the model.	Scenario based validation.	Explainability of method's decision discussed in Section 5.3.1 of Experiment 3 along with the risk assignment on transaction level.
Usability [C1]	EER connected with the FRR (False Rejection Rate), along with showcasing that the method can be used in financial application.	$\leq$ 1% EER. Scenario based validation.	Achieved EER validated in application design of Section 5.7. Possibilities of risk based authentication increasing usability presented in Section 5.8.1.
Privacy [A2]	Choosing features which are not privacy threatening. Not storing privacy threatening data.	Binary (Yes or No), proved using a data model that does not contain sensitive information.	Privacy preserving data model presented in Section 5.8.2, relying on the features used in data processing presented in Section 4.4.
Interoperability [P1]	Scenarios of use.	Scenario based validation.	Scenarios of deployment along with the data model showcased in Section 5.8.2.

Source: own elaboration

2. **Datasets** - method works based on already collected touchscreen data, in the form of datasets containing touch patterns of multiple users. The data are processed in the same way it would be when captured by the mobile OS API. The method will be verified on multiple datasets[1] available in the literature, including Touchalytics (Frank et al., 2012), Sapienta BioIdent (Antal et al., 2014), Serwadda BTAS dataset (Serwadda et al., 2013) - all described in the Table C and own dataset collected - described in Section 5.2.1. The preprocessing performed is described in Section 5.2.2. In terms of additional fraud detection possibilities, also BrainRun (Papamichail et al., 2019) dataset is used and described in Section 5.5. All of the underlying swipes' dataset formats will be unified by extracting the features from the Table 4.2, as long as the dataset contains information about a given certain feature group or sensor.
3. **Preprocessing** - due to the dynamic nature of touchscreen data a small part of falsely registered swipe events is filtered out (reference in Section 5.2.2). Removal of very short swipes (spanning over milliseconds) was also performed due to the fact that calculation of acceleration and speed values requires multiple readings. The feature values are normalized using min-max scaling and depending on it achieving higher result metrics compared to the non-normalized data, better score is always presented. Additionally, random sample of impostor data is chosen to supplement each user's data in a way described in Section 4.3.
4. **Learning and testing** - method uses machine learning classifiers employing 5 fold cross-validation and train/test split of 80/20. Algorithms used will include Python implemented: XGBoost, SVM and Random Forest. Each of the classifiers is enriched with a grid of hyperparameters and grid search is employed to find optimal accuracy and error rates.
5. **Optimization and evaluation** - the best classifier from the ones evaluated in the grid search procedure is chosen based on the evaluation metrics described in 4.5.1. The best model will then be evaluated in terms of meeting the requirements criteria, namely the EER in accordance with the [A1] and [C1] criteria presented in the Table 4.1. The resulting error metrics achieved on all of the datasets will be compared for 1,3,5 and 7 consecutive actions to achieve satisfying error rates. To confirm that the results can find unique

[1]While Touchalytics, Sapienta and BrainRun datasets are widely available, Serwadda BTAS was acquired from TouchDB benchmark at `http://atvs.ii.uam.es/atvs/TouchDB_Benchmark.html` based on signed license agreement.

patterns in all of the datasets and find whether similar patterns are observed despite the device used a comparison of feature importance are employed and presented in Section 5.2.3. This aims at providing the consistency, and generality of the developed artifact, in accordance with evaluation criteria presented in the Table 1.4.

6. **Fraud detection** - the developed method enables risk assessment for each transaction performed in the application, based on one or multiple actions performed by the user. The model meeting the fraud detection [B1] criteria is presented in Section 5.3.1. Further possibilities of automatically classifying other helpful information about the user preventing unauthorized access such as gender and age to enrich fraud detection systems are verified in Section 5.5.
7. **Architecture** - method's data model should be feasible to be deployed in financial services infrastructure and enable risk assessment in the application. The proof of meeting the Usability criteria in adaptive risk based authentication, relying on achieved error rates is presented in validation scenarios. Also in accordance with the interoperability [P1] and privacy [A2] requirements the data model and feasibility of implementation will be presented in Section 5.7.

The method is a decision based authentication which provides information about the risk of authorized access. It is aimed at classifying user activity as either valid or anomalous after multiple actions (mostly swipes). The potential of method's extensions to utilize taps instead of swipes will also be presented in Section 5.4. That way, based their behavioural profile, each user action can be measured in terms of similarity to the typical user behaviour. This allows for decreasing the potential need for interaction in terms of low-risk transactions and ensures higher level of security for high-risk ones. Additional benefit is the possibility of local risk based assertion of actions e.g., differentiating between allowing the user to access the account balance and making large money transfers. This local system of risk assertion can be connected with the already existing fraud detection systems.

4.3 Assumptions

The method's goal is to authenticate a user, there are however two approaches to this problem, both of which are described in the Figure 4.1:

- **1-to-Many** (identification) - simplifying the issue to a multi class classification problem, where based on supplied data, we try to predict which user is using the phone out of the collection of many. This description brings it to one of essential machine learning problems. Most of the approaches mentioned in the literature handle this issue as such. The identification approach however has a few issues. Firstly, it is susceptible to class imbalance problem, especially considering the datasets which did not follow a well defined scenarios, but rather just observed user activity. As differences may appear between the number of samples for each user, the algorithm may ”favour” a user with more samples as it would classify more samples correctly overall. Secondly, the scenario presented is unrealistic as it assumes identification of a user (based on X samples, classify which user it is) instead of an authentication problem. Thirdly, it can cause performance and accuracy errors, as each sample needs to be compared to an N number of users, which is ineffective in both the learning process outcome (due to multitude of classes and small number of samples) and the time required for the classifier to learn - as it requires all of the other users data to be included in the training phase for the recognition of a single user.
- **1-to-1** (authentication) - where we assume that only two classes in the classification scenario exist. As we label rightful owner’s actions as one class and assign all other user samples to the second one. This method of classification is closer to a real use case scenario for the method. It may however create even greater class imbalance problem, widely known in machine learning research. If we use too much other user’s data for the classifier, the training process will be long and not guaranteed to achieve better results. Due to that, we need to come up with a ratio of user and imposter data that is not so imbalanced and still allows the algorithm to differentiate the users well.

To comply with the literature often applying the identification approach, this work presents the observed performance differences (measured by accuracy metrics and error rates) achieved by comparing the same method in both of these scenarios. To combat the issue of class imbalance a sampling method could be used for the learning process - which could influence the classifier results based on similarity of samples to the original user. Similarly, some of the classifiers can be parametrized to combat this issue independently of sampling, by changing the weights of data labelled with a specific class. In turn, the learning error can be multiplied

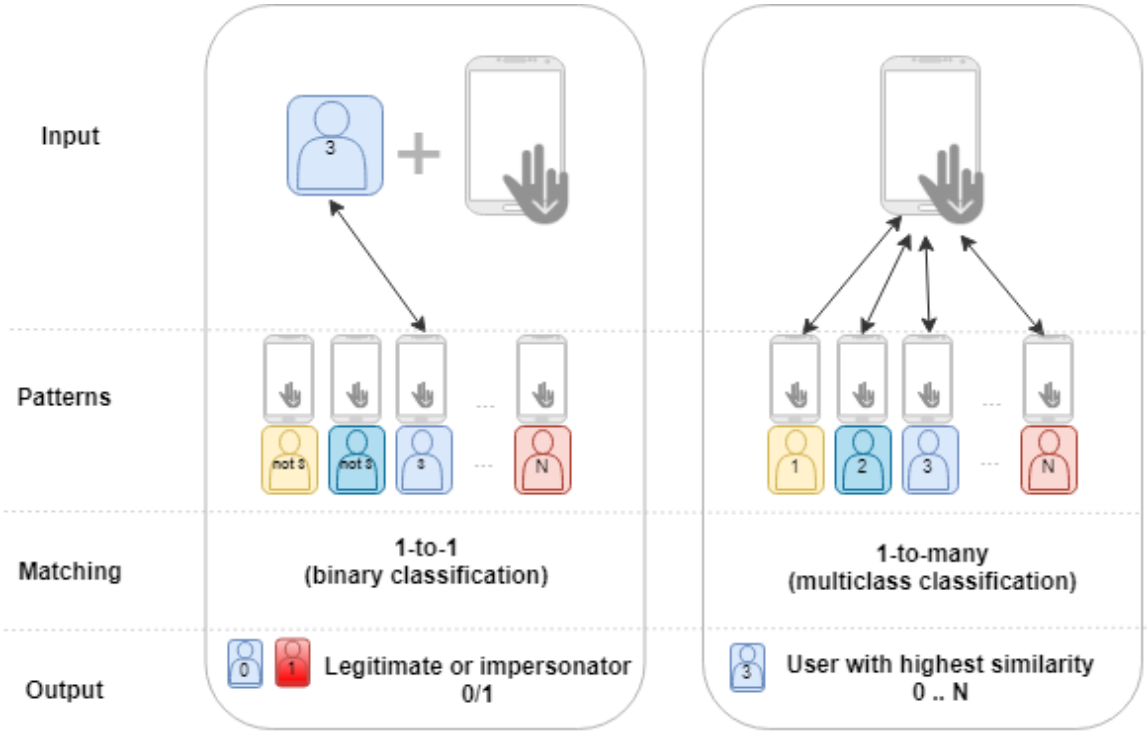


Figure 4.1. Differences between identification and verification scenario. Source: own development based on (Teh et al., 2016)

based on the inverse of the ratio of the class size. To perform the classification in an authentication scenario multiple, separate classifiers will be used for each user[2].

Inline with the above, the method employs 1:1 and 1:2 representation of user vs. impostor data in the authentication scenario to evaluate its results. This should not cause the classifier's errors to raise given the larger number of users in different datasets and is as close to the verification scenario of other biometric methods as possible[3].

There are also some other assumptions connected with the method that we need to place before the design and evaluation can begin. These assumptions are needed to provide a clear evaluation scenario, where we need to assume a certain level of stability for the environment in which we hope to perform our experiments on the datasets further on. They includes the following:

[2]What is worth to note is that this is significantly different from one-class classifiers which would assume only the user data is available and we do not have other users samples. This issue is described at the end of the Chapter 5.

[3]One-vs-rest classifiers could also be used, but they yielded slightly worse results in each case, than just labelling the data as 0 and 1 and employing sampling in the authentication experiments.

1. We assume only one user is using the device during the learning phase. We do not cover the issue of multiple users sharing the device, each non-matching pattern will be recognized as a fraudulent behaviour.
2. We do not aim to provide 100% secure authentication method which may rival fingerprint detection, but an inherence factor method which may be used inline with knowledge (passwords) or other factors in authentication process. Use of the method in multi-factor authentication may either increase the security or usability (as we may not require a knowledge factor if the method recognizes a user) and the financial service provided may shift this focus based on those criteria importance in a specific situation. Goal is to provide an alternative to mobile face recognition, not burdened with privacy or pattern leak problems that can also give additional information to fraud detection systems and possibly allow adaptive, continuous authentication.
3. We divide the method development between the design part (described in this chapter), evaluation part and the validation considering its use in the mobile financial applications. First part is where the method is created, second is where its performance and consistency is tested over multiple datasets. This is inline with the previous examples of behavioural biometrics literature evaluation methods. Last part is the scenario-based validation part where the requirements considering the privacy, usability, fraud detection and interoperability are shown to be fulfilled by the method's implementation scenarios.
4. We focus on touchscreen based authentication, inline with the findings from the previous chapter. This is due to their compliance with the requirements for the method described in Chapter 2. This may however include improvements to the current touchscreen biometrics methods in terms of enriching the interaction data with another sensors and features or utilizing ensemble models if necessary.
5. One class classification problem is not handled by the method - meaning the assumption that only the actions of genuine users are available as learning data. There are however approaches in literature (some as early as 2007 (Mazhelis, 2007)) which show the possibility of handling this issue with existing methods, without fundamental changes required. Such example for accelerometer data and gait recognition (on 4 datasets) - with results showcasing achieved accuracy metrics similar to multi-class approaches was presented by Kumar et al. in 2018 (Kumar et al., 2018). Similar approaches for touch interaction were proposed by Choi et al. and Yang et al. (Choi et al., 2018; Yang et al.,

2019). The approaches do not show significant drop in performance, assuming the distribution of the data among population can be studied, which should be the case before deploying it in an financial application.

4.4 Data processing and features extraction

The authentication approach used in this work assumes that the data from touchscreen sensors is collected during the interaction with a specific application. The features extracted from the system API were described previously in the Section 3.4.2. Based on this rough description of basic groups of variables, this work lists the possible features that may be extracted from specific measurements of the input sensors. They all may characterize the user movement when swipe is performed in the test application.

One of the features used by Frank et al. (Frank et al., 2012) was not used in the classifiers, namely the *inter-stroke time*. The definition of this feature, is that it measures the duration between the end of the swipe and beginning of the next one. This in turn decreases the number of samples in each user session by one (as the last swipe may not have this information) and makes the swipes dependent on each other in a sequential manner - which even if it influences the classifier positively may bias the results achieved. While for the well defined experiment or application use scenario (for example where the user requires N clicks to finish the process) this might be a useful feature, its use is not encouraged when comparing different datasets, varying tasks and session lengths.

The summary of features extracted from touchscreen and other accompanying sensors is presented in the Table 4.2. Each feature is numbered and assigned to a feature group which connects it to the categories mentioned in Section 3.4.2. The calculation of said features from the initially captured data follows the principles stated below:

1. The initial data consists of raw sensor input and low level system API information, which can be retrieved about any movement on the device[4]. This includes X and Y coordinates of the touch action, finger orientation, timestamp (represented as milliseconds in Unix time format) pressure and area of touch measured (that may be also expressed by minor and major ellipses of touch for some devices[5]). Additionally accelerometer and gyro-

[4] https://developer.android.com/reference/android/view/MotionEvent

[5] In Android API the respective functions are https://developer.android.com/reference/android/view/MotionEvent.PointerCoords#touchMinor and

scope readings in X, Y and Z axis can be joined with this data based on the timestamp of measurement. The touch *event* type is extracted to differentiate between the swipes, as ActionDown (moment when user begins the gesture), ActionMove, and ActionUp are extracted. Those values are available on both Android and iOS APIs, with different names as they are necessary for registering any kind of a gesture. An example of data in this raw format from Touchalytics dataset is presented in the Figure 4.2. The dataset contains additional information that identifies the user and a unique session it was captured in.

2. Based on a list of consecutive rows from the data format mentioned above, the swipes are extracted. All rows which **begin with ActionDown**, in most of the datasets identified as action with value 0, up to the closest in time **ActionUp (end of touch for a single gesture)**, identified as value 1, are taken as a slice of the dataset.
3. In this slice the first row is the beginning action of a swipe, and the last row is its end. From this information the basic statistics about the start point, end point, length (number of measurements) and overall duration/time of the swipe is extracted.

Next, when the collection of consecutive measurements corresponding to the single gesture, previously referred to as slice, is extracted, the feature calculation begins, covering variables listed in the Table 4.2.

To calculate the distance, speed (velocity) and acceleration, the consecutive measurements corresponding to a single swipe are used for extracting the difference in position in a Euclidean distance, over the time difference observed. That way in the data slice S describing a single swipe can be represented a n long vector of measurements s, where each measurement s_n contains the basic information: $s_n = [xcoordinate_n, ycoordinate_n, time_n, area_n, pressure_n ...]$ as explained previously. All of the measurements then are aggregated into the vector S in a following way:

$$S = [s_1, s_2 ... s_n] \tag{4.1}$$

The measurements in the swipe are sorted ascending by time. Having the information about the position of an action (expressed by its x and y coordinates) the speed (velocity) of a mea-

https://developer.android.com/reference/android/view/MotionEvent.PointerCoords#toolMajor

	phone_id	user_id	session_id	time	action	phone_orientation	x-coordinate	y-coordinate	pressure	area_covered	finger_orientation
0	0	36	3	1334893336544	0	1	272	269	0.21	0.04	0.00
1	0	36	3	1334893336790	2	1	262	271	0.32	0.04	0.00
2	0	36	3	1334893336795	2	1	123	327	0.28	0.04	0.00
3	0	36	3	1334893336800	1	1	123	327	0.28	0.04	0.00
4	0	36	3	1334893336885	0	1	216	298	0.34	0.04	0.00

Figure 4.2. Raw data from Touchalytics dataset. Source: own elaboration

surement s_n is defined as follows:

$$Velocity_n = \frac{\sqrt{(x_n - x_{n-1})^2 + (y_n - y_{n-1}))^2}}{t_n - t_{n-1}} \tag{4.2}$$

Where the Euclidean distance between the points in x and y screen coordinates is: $\sqrt{(x_n - x_{n-1})^2 + (y_n - y_{n-1}))^2}$, can also be expressed as d and the time difference $\Delta t = t_n - t_{n-1}$. Each of those measurements can be also represented as a vector - D for distances and T for time differences, and both are of $n-1$ length. The values can also be calculated separately for each axis - corresponding to rough approximations of horizontal and vertical velocity between each measurement.

$$\begin{aligned} XVelocity_n &= \frac{\sqrt{(x_n - x_{n-1})^2}}{t_n - t_{n-1}} \\ YVelocity_n &= \frac{\sqrt{(y_n - y_{n-1})^2}}{t_n - t_{n-1}} \end{aligned} \tag{4.3}$$

Further on, the acceleration can be calculated on the whole dataset as the difference of the speed observed on each measurement:

$$Acceleration_n = Velocity_n - Velocity_{n-1} \tag{4.4}$$

As the speed and velocity can be calculated between the measurement pairs, in the end we obtain a Velocity vector V with $n-1$ length, consisting of velocity values $v_1, v_2 ... v_{n-1}$ and an acceleration vector A with $n-2$ length which acceleration values $a_1, a_2 ... a_{n-2}$.

Analyzing the features presented in the Table 4.2, first group [1]-[4] represents the exact position (in pixels) of the beginning and ending of the gesture. As represented in the Figure 4.3 by s1 and s9 drawn as large black dots. Further on, the length [5] represents the number of samples - effectively n (9 in the Figure provided as an example). Time feature [6] is the difference between the last and first timestamp $time = t_n - t_1$, where n is the index of the last measurement. The **line distance** [7] is a straight line distance between the start point and

Table 4.2. Features identified for swipes classification.

Feature group	Feature name	Description
Positional	[1] start_x, [2] start_y, [3] end_x, [4] end_y	Position in pixels of the start and end points of the stroke.
Time and duration	[5] length, [6] time	[5] number of measurements in the swipe. [6] duration of the stroke in miliseconds.
Positional (all measurements)	[7] line_dist, [8] real_dist, [9] distance_ratio	Linear distance calculates the difference between the point (start_x, start_y) and (end_x, end_y) in Euclidean distance. Real distance is the sum of distances calculated for every consecutive measurement.
Positional	[10] xrange, [11] yrange	Maximal difference in X and Y axis.
Directional	[12] direction_atan, [13] direction, [14] hor_vert	[12] The atan2 value of the measurement defining the direction of a swipe. [13] The direction is then represented in 4 classes 1 - up, 2- left, 3- down, 4 right. [14] Horizontal/vertical stroke identification.
Directional (all measurements)	[15] mean_res_len, [16] mean_angle	[15] Mean Resutlant Length circ_r(atan2) [16] Mean angle (in arctan2 representation) of consequtive measurements for the whole stroke.
Positional (all measurements)	[17] largest_deviation, [18,19,20] q1_dev, q2_dev, q3_dev,	[17] Largest deviation from end-to-end line. Meaning the largest Euclidean distance observed for the whole swipe calculated from the line going from StartPoint to EndPoint directly. [18,19,20] Respective quartiles for the same vector of distances from straight line.
Speed and Velocity (all measurements)	[21] s_beg, [22] s_end, [23] s_mean, [24] s_iqr	Speed (calculated in pixels/milliseconds) for the beginning of the stroke, end, mean speed and Inter Quartile Range (IQR) of the speed.
Speed and Velocity (all measurements)	[25] a_beg, [26] a_end, [27] a_mean, [28] a_iqr	IQR and mean values of the acceleration vector (calculated in pixels/milliseconds) during the beginning, end and the whole gesture.
Speed and Velocity (all measurements)	[29, 30, 31, 32, 33] x_speed_b, x_speed_e, x_speed_mean, x_speedrange, x_speed_iqr, [34, 35, 36, 37, 38] y_speed_b, y_speed_e, y_speed_mean, y_speedrange, y_speed_iqr	Speed values (beginning, end, mean, range and IQR) for x and y axis separately, similar to [21-24].
Pressure (all measurements)	[39, 40, 41, 42] pressure_mean, pressure_median, pressure_iqr, pressure_range	Pressure measurement average, median values and IQR and range. Quartiles are not calculated due to low variability of the feature.
Touch size (all measurements)	[43, 44, 45, 46] size_mean, size_median, size_iqr, size_range	Size (based no TouchMajor and TouchMinor ellipses) measurement average, median values and IQR and range. Quartiles are not calculated due to low variability of the feature.
Finger and device orientation	[47, 48] finger_orient, finger_orient_change, [49, 50] orientation,orientation_change	Finger and device orientation at the start of the stroke [47,49] and range of value changes [48,50].
Pressure / touch size	[51, 52, 53] mid_press, mid_area, mid_finger_or	Calculated values of pressure, size and finger orientation for the middle of the stroke.
Accelerometer (all measurements)	[54], [55], [56], [57] all_dev, dev_avg, dev_iqr, dev_range, [58], [59], [60], [61] x_dev, x_avg, x_iqr, x_range, [62], [63], [64], [65] y_dev, y_avg, y_iqr, y_range, [66], [67], [68], [69] z_dev, z_avg, z_iqr, z_range, [70], [71] cor_xz, cor_yz	The x [58-61] and y [62-65] and z [66-69] axis readings of the accelerometer. With the resulting overall movement of the accelerometer defined as $dev = \sqrt{(X^2 + Y^2 + Z^2)}$, and the respective descriptive statistics are also calculated for the values of deviations (dev). Also as proposed by (Singha et al., 2017) the respective ratio of average X [70] and Y [71] to Z axis is calculated.
Touch size minor and major ellipse	[72, 73] minMajor, maxMajor, [74, 75] minMinor, maxMinor, [76, 77, 78] major_avg, major_iqr, major_range, [79, 80, 81] minor_avg, minor_iqr, minor_range	Due to the potential of using minor and major axis of the ellipse for the touch event in the Android API, those variables could also be extracted to improve the classification accuracy.

Source: own elaboration

the end point: $\sqrt{(x_n - x_1)^2 + (y_n - y_1))^2}$, where the real distance d [8] is the sum of collective distance measurements d_i in the sample:

$$d_i = \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1}))^2}$$
$$real\_dist = \sum_{i=0}^{i=n} d_i \tag{4.5}$$

And the distance ratio [9] is $\frac{line\_dist}{real\_dist}$. The distance ratio of 1 means the user exactly followed the shortest path to the point, where the smaller the value the more broad the movement of the user was. The ranges of x [10] and y [11] represent the maximal difference in those axes as:

$$xrange = \max(x) - \min(x)$$
$$yrange = \max(y) - \min(y) \tag{4.6}$$

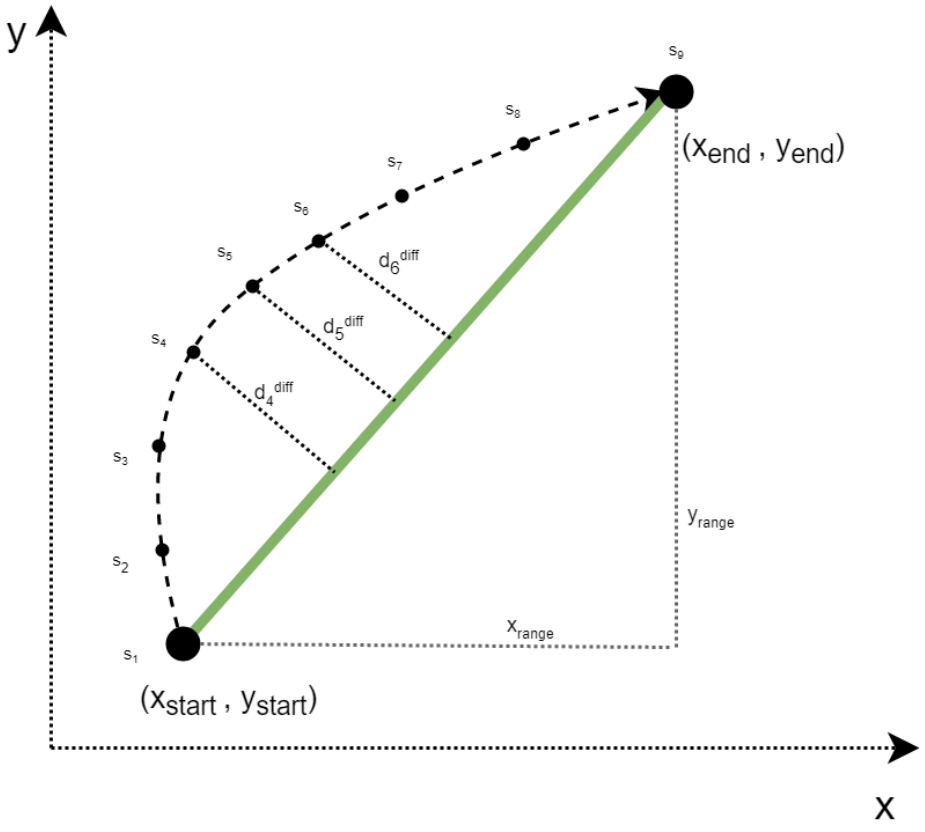


Figure 4.3. The representation of a swipe with a real distance calculation. Source: own elaboration

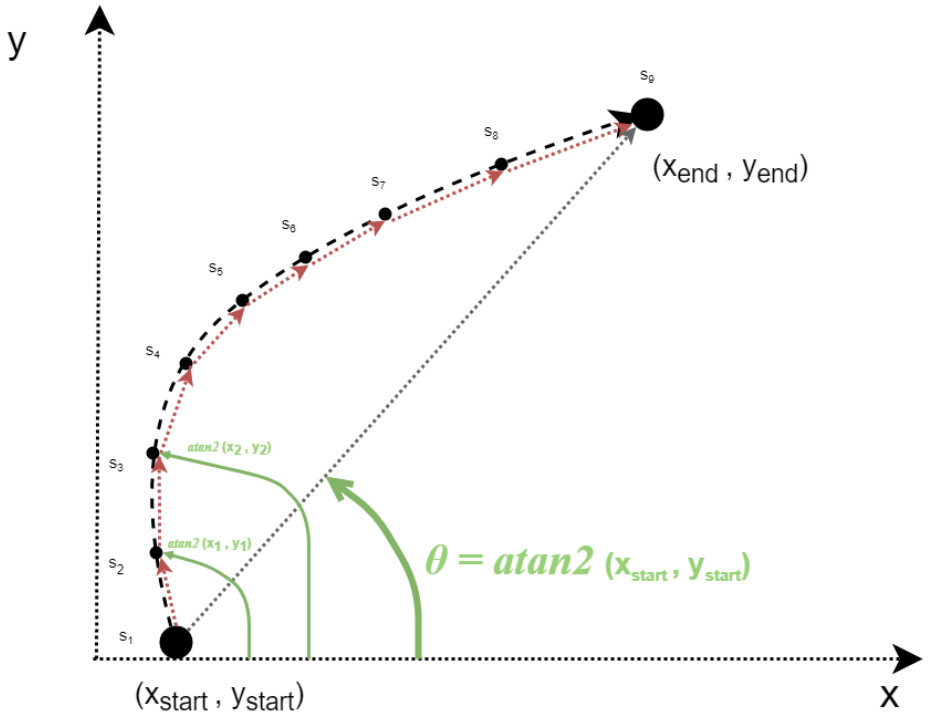


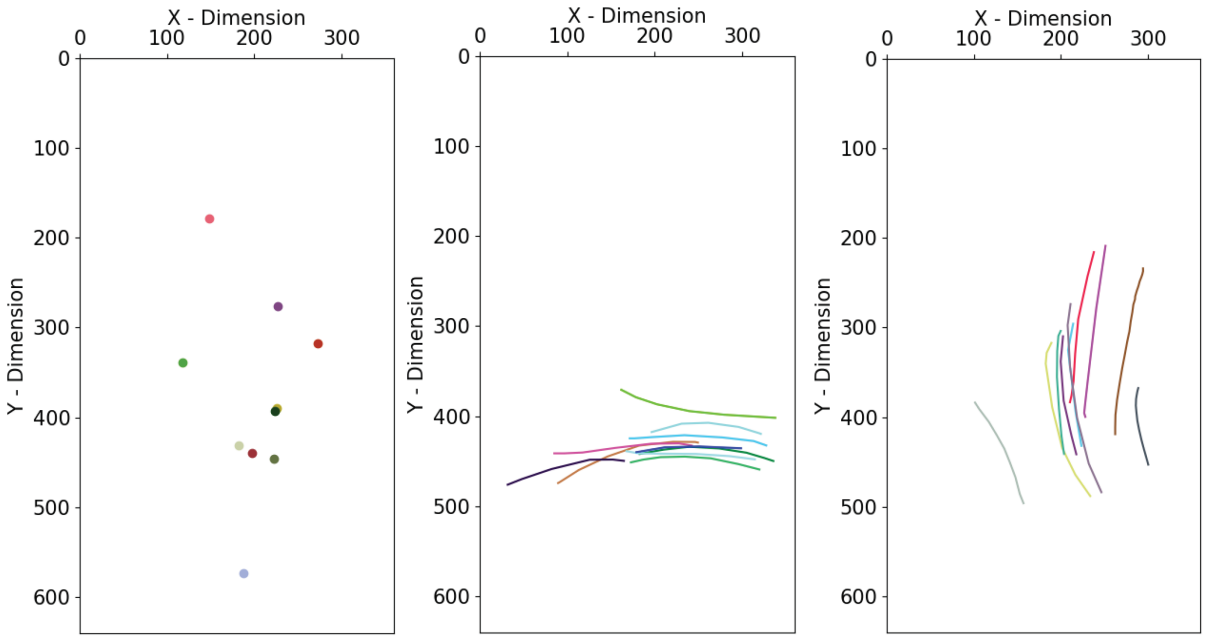
Figure 4.4. The representation of a direction calculation for a swipe gesture. Source: own elaboration

The direction of the movement and angle is first calculated based on the angle the movement is taking place, considering the standard vertical position of the phone[6]. The direction of the whole swipe [12] is calculated based on the following function $\Theta = math.atan2(end_y - start_y, end_x - start_x)$. Then the swipes are classified to one of the four directions: up, left, down, right [13]. Based on the output of the function the swipes are then classified as either horizontal or vertical [14] based on the output of the function. The different ways users perform these swipes can be seen in the Figure 4.5.

Next, the mean resultant angle is calculated [16] as a simple mean observed radian angles between the consecutive measurements. For the calculation of the concentration of angle the circ_r function is used[7] which calculates [15] the length of the mean resultant vector and is often used in hypothesis testing in directional statistics.

[6]Where the data was available the position of the phone was also included as a ”phone orientation” feature. This made it possible for the classifier to handle direction uniquely for those two modes of use.

[7]https://pingouin-stats.org/generated/pingouin.circ_r.html



(a) Taps (b) Horizontal swipes (c) Vertical swipes

Figure 4.5. Examples of different taps and swipes for users. Source: (Papamichail et al., 2019)

Further on, the distance between the straight line and observed movement is calculated for each point and added to the deviation vector.

$$D_{diff} = [d_1^{diff}, d_2^{diff}, ...d_n^{diff}]$$
$$d_n^{diff} = d(s_n, Y_{line}) \tag{4.7}$$

where n is the position of given sample (1 for S1, 2 for S2, etc.) and the task is to calculate perpendicular distance from s_n to the straight line passing through the start and end points - Y_{line}. This line is shown in bold green in the Figure 4.3 and the representation of d_n^{diff} distances is shown directly as dotted lines. The maximal value of this vector is then calculated as a largest deviation from the straight line that was observed in the gesture [17].

The variables 18 - 46 are calculated in a similar way. For each of the vectors (D_{diff}, *Velocity*, *Acceleration*, *XVelocity*, *YVelocity*, *Pressure*, *Size*) statistics can be calculated which emphasize their uniqueness. The mean, range and median is also calculated for most of the features.

The quartiles - Q1, Q2 and Q3 are calculated and the value of inter quartile range (IQR) is also calculated as:

$$IQR = Q3 - Q1 \tag{4.8}$$

This covers the features [18-20] for the deviations from the straight line d^{diff}, for the velocity vector IQR [24], [33] and [38] for the velocity in X and Y axes and for the acceleration [28]. The same approach is employed for the observations of touch pressure [41] and size [45].

For all of those basic values, based on the empirical observations of (Frank et al., 2012) feature importances while recognizing users, the values of of velocity at the beginning and the end of gesture proved to be of high importance to the classifier (19,63% and 7,85% respectively). Following this finding, the following heuristic is used for classifying the number of measurements of each gesture:

```
perc_25_index = (length-1)
#  5 or less points from the beginning or an end
beg_end_index = min(5, perc_25_index)

#  beginning vector
vector_b = vector[:beg_end_index]

#  end vector
vector_e = vector[-beg_end_index:]
```

Listing 4.1. Beginning and end of swipe indexing.

This index value is calculated in a way that it works even for short strokes. And the _b and _beg indexed features utilize the first values up to the index and the _e and _end cover the last index number of observations.

The index of the middle part of the stroke is used to characterize the movement also, providing the value of pressure, area and finger orientation observed in the middle observation [51-53].

For the accelerometer readings, as they cover the axes *X*, *Y* and *Z* axes as shown in the Figure 4.6. For these axes the overall deviation is calculated:

$$dev = \sqrt{(x)^2 + (y)^2 + (x)^2} \tag{4.9}$$

The mean, IQR and range of these values is calculated. As observed by the (Singha et al., 2017), the Z axis mostly remains stable, hence the ratio of average correlation of XZ and YZ axes may be a unique feature identifying a user, which is calculated as features [70-71]. There is still possibility of utilizing gyroscope readings statistics similarly to the accelerometer, but none of the datasets used included this type of information.

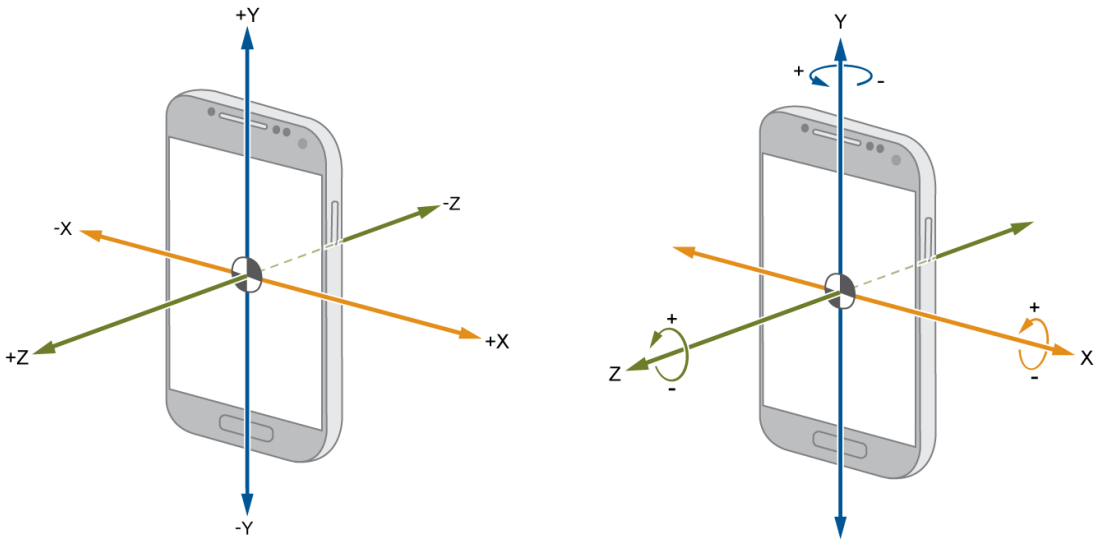


Figure 4.6. Phone accelerometer (left) and gyroscope (right) measurements in the 3 axes: X, Y, Z. Source: (www.mathworks.com, 2020)

Additionally, the values of the minor and major axes of touch area observed are calculated[8], along with their min, max, mean, range and IQR [72-81].

4.5 Classifiers used

For the user identification and authentication the following classifiers were used on the data prepared as described in the previous section:

- SVM - implemented by ScikitLearn Python svm module[9],
- Random Forest (RF) - implemented in ScikitLearn Python RandomForestClassifier module[10],
- XGBoost - implemented in the xgboost Python library[11].

Widely used algorithms, such as SVM and RF, have been proven to provide SotA results in touchscreen behavioural biometrics classification (Frank et al., 2012; Patel et al., 2016; Syed et al., 2019). To add to those methods and possibly improve the achieved accuracy, a novel

[8]API documentation: https://developer.android.com/reference/android/view/MotionEvent.PointerCoords.html#touchMajor
[9]https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html
[10]https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[11]https://xgboost.readthedocs.io/en/latest/python/python_intro.html

XGBoost algorithm described in 2016 by (T. Chen & Guestrin, 2016) was used. It has been proven to provide results better than standard Random Forest algorithm, be more resistant to overfitting and faster than other Python-implemented classifiers. It also deals quite well with under and over sampling due to the option of automatic sample weighting in the dataset.

The classifiers are to be used to train a model which will recognize user's actions - swipes in all experiments besides Experiment 4 5.4, where taps are used instead. In classification scenario:

- Experiment 1 (Section 5.1) and 2 (Section 5.2) employ the classifier to predict to which user the sample belongs to in an identification scenario.
- In terms of authentication scenario from Experiment 3 5.3 the classifier is supplied two classes: 0 (user) and 1 (random impostor) and the algorithm should predict whether unauthorized access took place along with the probability value of sample belonging to the rightful user.

All of the classifiers were up to date versions in 2020, where most of the experiments were conducted, downloaded by pip Python package manager.

The data will be split in 80% - 20% train/test subsets by train_test_split method of sklearn.model_selection and their learning process was performed by the Python GridSearchCV[12] procedure with cross validation parameter set to 5. The values of hyperparameters used for optimization are provided in the experiments. While the first experiment used significantly larger parameter range (Table 5.1), authentication experiment however utilizes smaller range of hyperparameters (Listing 5.2) to achieve performance increase.

4.5.1 Optimization criteria for the learning process

The machine learning algorithms that are to be used in the evaluation of the proposed method require a choice of criteria for error minimization. These criteria can then be used in the training and test phase of the algorithm to hopefully achieve the lowest error rates, corresponding to the minimal value achieved by the error criterion during the learning process. Observing the plot of these functions during this process will also help in uncovering potential errors and unwanted interdependencies between the predicted variable and predictors. Two loss functions were used in this work:

[12] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

- **Logarithmic loss** in classification measures the performance of a model where the prediction input is a probability between 0 and 1. It is defined as the negative log-likelihood of the true labels given a probabilistic classifier's predictions. A goal of the classification algorithm is to minimize this value. A perfect model would have a log loss of 0. Predicting a low probability (0,1) where the label is truly 1, results in a high log loss value, which the algorithm tries to optimize. The function is a logarithmic representation of how sure the algorithm is of the predicted class. As the predicted probability approaches 1, log loss slowly decreases. Log loss penalizes both types of errors, but especially those predictions that are confident and wrong - which makes it good in our case to penalize FAR metrics. Log Loss takes into account the uncertainty of your prediction based on how much it varies from the actual label. To evaluate a model against a dataset, the loss score will be an average log loss across all observations. To explain this metric even more, having:
  - N - number of observations,
  - M - number of possible class labels (user1, user2 or 0-owner, 1-impostor),
  - log - the natural logarithm,
  - y - a binary indicator (0 or 1) of whether class label c is the correct classification for observation o,
  - p - the model's predicted probability that observation o is of class c.

  The function is defined as:

$$\sum_{c=1}^{M} = y_{o,c} \log(p_{o,c}) \tag{4.10}$$

  where the $\sum$ is the sum of all log loss values across classes c from the first class (c=1) to the last classes. In authentication scenario, relying only on two classes of user (0) and impostor (1) only one class is then calculated.
- **Multi-class classification error rate**. This simple metric is calculated as $\frac{\#wrongcases}{\#allcases}$ and showcases the overall portion of wrongly classified users. Is is easy o interpret and its visualization and observed oscillations during the learning phase may showcase the method stability. The criterion itself applies also for an authentication scenario with only two classes.

4.5.2 Hyperparameter optimization strategy

Important part of the used algorithm is the choice of the **hyperparameters**, which define how the method classification is bounded and influence the computational complexity.

For the **Support Vector Machine** (SVM) classifier the kernel function, gamma and C are used. Explaining definitions of those hyperparameters is not crucial for this work, as there is a wide variety of sources which provide them[13]. What is important however is the influence of those parameters on the model. First of all, the kernel highly influences the results of the classifier as non-linear kernels tend to work better with noisy data. While rbf (radial basis function) and polynomial kernels may be more computationally complex, they were proven to work well with this type of data (which comes up to the variance in the data distribution, sometimes linear kernel may be enough if the data is generated by a predictable process). The C parameter controls the trade off between smooth decision boundary, as smaller values may generalize better (lower bias, high variance) and high C values may tend to the algorithm overfitting on the train dataset (higher bias, low variance). Gamma is a parameter tied to the influence of a single observation on the decision boundary. Higher gamma will lead to higher bias and low variance. There is also a degree parameter for the polynomial kernel, which in the end influences how the algorithm handles specific non-linear connections between the variables.

For the **Random Forest** (RF) algorithm, first the criterion used for defining the splits in the data is chosen in the bootstraping of trees. The bootstrapping can also be turned on or off by the bootstrap parameter. It is effectively a randomization technique that samples the data used to build a single classification tree. This in turn may make the classification accuracy better for some edge cases where similar observations may appear in the data, which belong to the different classes. Maximum depth (max_depth), max number of features used in one tree (max_features), minimal samples needed to create a classifying leaf (min_samples_leaf) or a split (min_samples_split) are tied to a singular tree classifier. Those parameters put boundaries to the complexity of the tree, high maximum depth and low minimal number of samples may lead to overfitting, while the opposite leads to low accuracy of the classifier. The number of estimators-trees used in the classification (n_estimators) is the most computationally demanding hyperparameter, as it effectively trains an *n* number of tree classifiers to form a

[13] https://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html, https://scikit-learn.org/stable/modules/grid_search.html, more information is available in (Kecman, 2005)

final decision. More classifiers means less possibility of overfitting but significantly increases computational complexity.

For the **XGBoost** algorithm, effectively used as a boosted RF classifier, similar parameters to the above mentioned model are used. This includes maximal depth, minimal leaf split (min_child_weight) and the number of estimators. Additionally the subsampling method is used, which takes 0,8 of input train data for building each tree. The eta parameter and learning rate are also used - which define how quick the algorithm is in creating rules based on the observed relationships in the data. High learning rate with the low number of estimators may make the algorithm work faster in a production environment, but low learning rate and high number of estimators may give a statement about the boundaries of the classifier's accuracy which can be achieved. The objective function used is chosen for the multiclass classification (in terms of classifying an observation to a single class from a representation of N classes) and the is_unbalance parameter automatically assigns weights to under-sampled classes (classes with the lower that average number of samples). No missing values are passed to the method, hence the parameter is set to None.

In order to determine the optimal values of parameters that maximize the classifier effectiveness and minimize the loss function used, hyperparameter tuning is used for a given model. This is significant, as the performance of the entire model is based on the hyperparameters values specified. For optimizing those parameters a standard grid search approach is used that will try all combination of parameters to come up with the best performing classifier. There are other approaches, such as randomized parameter optimization, but may be computationally costly and due to the much needed evaluation for multiple datasets and limited time - the grid search may provide similar values with much less computation time required. For the first experiments hyperparameter grids focused on achieving absolutely high accuracy are used, hence very large number of estimators and low learning rates. Later on, for the practical scenarios a more balanced approach is used that focuses on the practical possibilities of performing such optimization, which will influence the variables range in the grid.

4.6 Added method benefits over SotA

Compared to the 32 main features extracted by the Frank et al. (Frank et al., 2012) and the same number identified by (B. Zou & Li, 2018). This work proposes overall 81 features, consisting of:

- 22 more features extracted from basic positional and directional information about the swipe,
- 10 features containing information about the minor and major ellipse axes for the touch area in each point measured,
- 18 features extracted from the accelerometer readings if they are available to supplement the swipe data. While not collected in the datasets in the literature, they are included in the own dataset collected for the purpose of evaluating the combinations of features used. Similar variables could also be extracted from the gyroscope sensor, but it was not supplied in the dataset and can be the subject of further work.

The number of features identified in this work surpasses examples described in recently published reviews of the literature on similar datasets (Fierrez et al., 2018; Patel et al., 2016).

The method's approach uses XGBoost, which is compared to the often used Random Forest and SVM classifiers that are normally described in the literature. Although classification scenario is presented in both Experiments 1 and 2, the addition over SotA comes with the introduction of unified authentication scenario with 1:1 and 1:2 impostor data ratios. Although the approach is similar to the one class classification in the machine learning literature, this work provides unification of the evaluation process and describes the differences in achieved error rates between the evaluation processes and impostor data ratios, in turn extending the results achieved in the literature. The approach presented allows the classifier accuracy and computational requirements of model learning to not decrease if the number of users is larger in the dataset. Presented comparison of feature importance has not been done previously in the literature. It is aimed at answering which touch features contribute the most to the uniqueness of the user pattern and if they are uniform across multiple experiments, devices and datasets. Overall the proposed approach defines performance after a set number of actions - swipes, which allows the comparability of results over multiple datasets. Although using multiple actions to authenticate is not new (however, the number is not uniform across datasets, which this work provided), it has not been validated yet in the application design for the financial domain, which the work provides, along with the accompanying data model and architecture proposals.

4.7 Summary

The chapter presented the design of the main artifact of this dissertation - the touchscreen authentication method. The goal of the designed method is to authenticate users in the mobile environment of financial services. The developed method answers the **RG4** and fits into the model of requirements defined in Section 2.3.4. The procedure of validation of the method's performance required some assumptions about the method. They were presented in the first part of the chapter and outlined the potential issues which may be out of scope of the experiments presented further on. To meet the criteria for performance and error rates, two distinctive approaches to testing the method's accuracy were presented, relying on the identification and authentication schemes. The verification and validation scheme of the method according to the requirements model criteria was presented in the Table 4.1.

Further on the designed method was described, focusing on explaining the various features which will characterize a swipe for touchscreen authentication. The summary of said features was shown in the Table 4.2 and partially fulfilled the **RG3** in terms of providing a list of variables which may identify and authenticate users based on the chosen touchscreen behavioural biometrics modality. In the last part of the chapter, in Section 4.5 the classifiers that are used to train and evaluate the method were presented along with their hyperparameters and the optimization strategies that were employed. Further on, the dissertation aims at designing the experiments which prove the method's compliance with the criteria set for the evaluation and validation procedures.

Chapter 5

Evaluation and Validation

The goal of the chapter is to evaluate the method design created previously, to conform to the research rigour required by the Design Science principles. The aim of the evaluation procedure is to prove that the method, main artifact produced by this dissertation, meets the evaluation criteria which were presented as the general requirements at the beginning of the design process by RG1 and were specified in the Table 4.1. Measurement of obtained values in multiple experiments is aimed at producing scientifically viable results which can improve the knowledge base of the domain. All datasets used in experimentation are described in the appendices in the Table C. On the other hand, usefulness of the artifact and its validity for the environment in which it was created is covered by the validation procedure, which aims at presenting that the artifact meets the requirements of the environment (including financial institutions, customers and third parties) and can be a valid method for authentication.

For the evaluation criteria, the main focus of this chapter is centered around the activity dimension from the Table 1.4. In terms of the research goals, this chapter aims to answer:

- **RG3** - by analyzing the importance and stability of the touch features' significance across multiple datasets (Section 5.2),
- **RG5** - in terms of evaluating the designed artifact as a whole, especially focusing on the error rates connected with the authentication method and possible scenarios of use. These scenarios prove that the method can increase the usability (e.g., providing continuous authentication) and can be implemented in financial applications. Referred to each of the experiments and validation scenarios presented in this chapter along with the criteria it aims to meet.

5.1 Experiment 1 - classification, user identification scenario

- **Goal:** preliminary evaluation of the method in a classification scenario, to compare with SotA. Measuring the decrease in error metrics when using multiple actions.
- **Datasets used:** Touchalytics.
- **Criteria:** validation of the method's preliminary results correctness and evaluation of the designed approach to classification of multiple actions.
- **Description of the experiment**: Touchalytics data is preprocessed by the method to obtain features representing user swipes. Three classifiers evaluate the results based on grid search optimization with 5 fold cross-validation. The learning process results are presented to showcase overfitting was not observed and there were no errors during the implementation process. Finally, multiple actions are used for prediction of the final class.

For this experiment the Touchalytics dataset was used. The raw data format is presented in the Figure 4.2 and consists of **912 133 datapoints for 41 users over 6 sessions**. It includes phone id, user id, session id, timestamp, action (ActionDown, ActionMove, ActionUp), phone orientation, coordinates of the action, pressure and area readings. Based on the features identified in the Table 4.2, the dataset was transformed to a format, where one row represented a single swipe. Due to the lack of accelerometer and major/minor axes for touch area only 53 of the features could be extracted. An additional variable containing the ID of the session was also used, as it distinguished between the task set for the user, which could imply different touch patterns.

The resulting dataframe was then split between train and test datasets (the test data accounted for 20% of the main dataset). The validation dataset was not created, due to the dataset not being significantly large. Also the cross-validation procedure effectively performed validation from the training dataset sample. The model achieved about 82,36% accuracy in predicting the user based on 1 swipe data. From all of the methods XGBoost performed the best, compared to the SVM and RF. The best parameters found by the CV (cross-validation) procedure were 'gamma': 0.5, 'learning_rate': 0.02, 'max_depth': 30, 'min_child_weight': 4, 'subsample': 0.8. It is interesting that such a large maximal depth of the tree was found to be optimal, pointing out to possibly complex relationships between the variables. Nonetheless, 5 fold CV procedure along with the subsample parameter prevented overfitting. The results

achieved are described in the Table 5.1. In case of the SVM algorithm normalization of numerical features was necessary for it to finish in a reasonable time, due to the distance metrics calculations required internally by the classifier. The Random Forest (RF) and XGBoost classifiers used non-normalized data.

The problem, being a multi class classification issue, was evaluated by the use of two criteria: merror and mlogloss. Those exact criteria were chosen for this problem and aimed at observing if the inferences made by those models are not caused by any errors in the data and prevent creating an overfitted model with high n_estimators parameter. The figure presenting the progress of the XGBoost classifier error minimization is shown in the Figure 5.1. As it can be observed in the first figure, the algorithm learned quite effective up to about 200-300 estimators range. There are no signs of overfitting or unpredictable behaviour. The expected fluctuations in the error rates are also present in the merror curve. Out of the feature importances identified by the model (which are not showcased in detail here), start and end points coordinates have been found to be the most important, along with the mean of size and the pressure registered. Further results point out to the possibility that users are being characterized by the specific point in which they begin and end their swipes, along with the angle identified between the start and end point (direction_atan). As expected, pressure and size of the finger contributed greatly in differentiating between users.

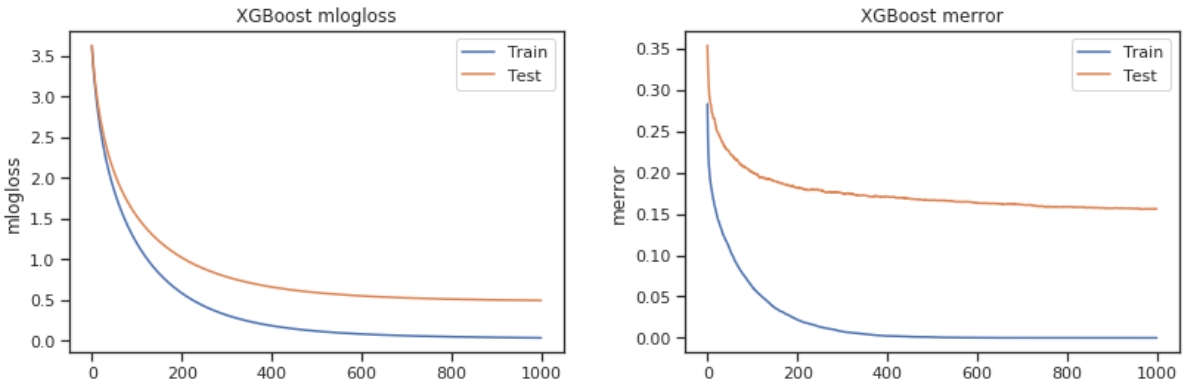


Figure 5.1. Error metrics for the learning curves of XGBoost classifier on Touchalytics dataset. Source: own elaboration

5.1.1 Multiple actions classification

The 82% accuracy achieved in the initial experiment is not enough for a reliable classifier in the financial environment. But the classification was made without accelerometer and gyro-

Table 5.1. Results achieved with the method on Touchalytics dataset. Train/test split 0,8/0,2.

Method	CV	Grid	Accuracy on test set
SVM	5	'kernel': ['linear', 'poly', 'rbf'], 'gamma': [0.001, 0.01, 0.1, 1, 10], 'C': [1, 10, 100, 1000], 'degree': [1,2,3,4] - only for 'poly' kernel	77,21%
Random Forest	5	'bootstrap': [True], 'max_depth': [None], 'max_features': ['auto'], 'min_samples_leaf': [1, 2], 'min_samples_split': [4, 2], 'n_estimators': [10,100,1000], 'criterion': ['gini', 'entropy'],	80,29%
XGBoost	5	max_depth=[15,30,45,60], gamma=[0.5, 1], min_child_weight=[2,4], subsample=[0.8, 1], learning_rate=[0.01,0,02], n_estimators=1000 eta = 0.1, objective = 'multi:softmax', is_unbalance= True, missing=None)	82,36%

Source: own elaboration

scope readings and additional data such as tap data was not used. What is the most important, and was missing in most of the studies up to date, excluding (Antal et al., 2015; Fierrez et al., 2018), is that the classification is performed often only after 1 swipe. This might be valid from a scientific and biometrics testing perspective, as is it the traditional way of evaluating such methods. In most of the cases this is a valid procedure. For example, in standard biometric authentication the user is mostly asked to perform a fingerprint scan once and there is no way of efficiently capturing it multiple times without a significant decrease of usability. However, in case of the touchscreen biometrics, the user is expected to perform multiple actions where these features can be extracted without hindering the usability of the process. In a real use case scenario user is expected to perform 3, 4, 5 etc. swipes before getting accessing an application function. While looking at an account balance may be available right from the start, performing a transfer requires multiple actions. To evaluate the classifier in such a scenario the following approach was applied, which consisted of:

1. Choosing a maximum number of strokes that are used in the classification, the maximum *n* value. *n* - 7 in this case.

2. Grouping the matrix (in the format of a python Pandas type dataframe) rows, representing swipes, by the unique *user_id*.
3. For each group, sorted by time, choosing only the ones with at least *n* samples in the test set.
4. Predict the resulting M rows, starting from i=0, chose M samples: $SAMPLE[i : i + n]$. If sample is longer, increase the *i*, up to the sample's length. For example in the case of 5 rows indexed as [1,2,3,4,5] and $n = 3$ the classification is carried out three times on [1,2,3], [2,3,4] and [3,4,5] slices of the dataset. This approach was basically a moving average over the prediction values of the classifier, which resulted in more results obtained over a randomized sample of test data.

The code used for grouping is provided in the Listing 5.1, relying on the use of Python Counter library[1].

```
y_col = 'user_id'

def multipleRowsClassification(group_rows, n):
    '''
    group_rows - the Pandas .groupby() object, representing user swipes
    n - number of rows based upon which the classification is performed
    y_col - is the column with user identifier

    Results:
    correct - TP + TN
    fail - FP + FN
    done - TP+TN+FP+FN
    accuracy = correct / done
    '''

    group_rows = group_rows.reset_index(drop=True)
    true_ylabel = str(group_rows[y_col][0])
    local_X = group_rows.drop(columns = [y_col])

    #value should not change in all N samples for the user
    assert group_rows.true_ylabel.nunique() == 1

    for i in range (len(group_rows) - (n - 1)):
        predicted_ylabels = chosen_model.predict(local_X[i:i+n])
        counter = collections.Counter(predicted_ylabels)
        prediction = str(counter.most_common(1)[0][0])

    global done
    done +=1
    if prediction == true_ylabel:
        global correct
        correct +=1
    else:
        global fail
        fail +=1
```

Listing 5.1. Grouping of multiple rows for classification.

[1] https://docs.python.org/3/library/collections.html#collections.Counter

The resulting best classifier was evaluated in that scenario, providing the following improvement in accuracy:

- after **3** swipes - **93.24%**,
- after **5** swipes - **98.28%**,
- after **7** swipes - **99.48%**.

The results achieved can be compared to the studies on similar data. Antal et al. from 2015 (Antal et al., 2015) achieved in the best case about 97,5% accuracy after 7 gestures for performing user identity classification scenario, and 70% from one swipe. Similarly, Fierrez et al. (Fierrez et al., 2018) achieved EER of <10% on a similar number of swipes. This might prove that the proposed method may achieve error rates satisfying the requirements, but comparison for multiple datasets is necessary to showcase that results are general and can be repeated if sufficient data is available, regardless of the dataset specifics.

5.2 Experiment 2 - comparison of accuracy on multiple datasets

- **Goal:** evaluation of the method in a classification scenario, to compare with SotA. Measuring the decrease in error metrics when using multiple actions.
- **Datasets used:** Touchalytics, Sapienta BioIdent, Serwadda BTAS, own dataset.
- **Criteria:** security [A1] - in terms of EER in classification scenario on multiple datasets to prove general nature of obtained results. Explanation of feature importance across the datasets aims to answer the **RG3**.
- **Description of the experiment:** data from all datasets is preprocessed by the method to obtain features characterizing swipes. Three classifiers are evaluated, and results are described using EER and other metrics such as precision, recall, f-score to provide insights into method's performance. Further on, possible metrics of feature importance are explained and the comparison of features is conducted on multiple datasets.

To evaluate the universality of the method's design and the approach for multiple actions classification, tests on multiple datasets had to be conducted. As such, the datasets found in the literature are used to showcase the method's accuracy. Unfortunately, accelerometer and gyroscope features mentioned in the Table 4.2 are not available for other datasets in the

literature. Due to that, as a supplement to the literature results, own dataset was created from a data-collecting Android application, enriching the data model with inertial sensors data. The main goal for the creation of this dataset was to showcase the potential benefit of using accelerometer and gyroscope readings and to test the method on users with different gender and age (as this data was rarely supplied by other datasets or all users were of similar age).

5.2.1 Own dataset

To compare the results achieved using different datasets in the literature and build on an existing knowledge about the features a test application was built. Its goal was to measure the tapping and scrolling gestures of users during the interaction and the accelerometer readings.

The data collecting application was built in a study supervised by the author and the co-promotor of this dissertation - Agata Filipowska. The code for the Android application was written by two 3rd year bachelor's students: Weronika Wąsowska and Paweł Wojciechowski. The study was performed between July 2019 and April 2020. The application design was quite simple and is presented in the Figure D.1. It was similar to the design of the Touchalytics and H-Mog applications from the literature. In the first screen (on the left) the users provided their: nick, age group (with the brackets for every 10 years of age, <10, <20, <30 etc.) and gender.

Next, the user is prompted to play a simple game, where he/she is asked to match the colours of symbols appearing on the screen. The user is to find all matching symbols. The idea is that a user utilizes two main gestures during the interaction with the application:

- taps - for clicking the specific symbol,
- scrolls - for moving the screen to access the next set of elements.

The representation of these actions during the interaction with the app can be seen in the Figure D.2. While we assume specific actions, all information about collected gestures is extracted by the Android API - when e.g., a user scrolls to select a symbol or tries horizontal scrolling. The data collected for the application covered:

- **88 unique users** - 46 female and 42 males. The users represented different age groups from 10-20, up to 60+ years. Age distribution of users and samples extracted is shown on the right side of Figure 5.6.
- **2009 unique swipes** extracted, **1497 unique taps**.

Table 5.2. Basic descriptive statistics for the dataset collected in the application.

Statistic	min	Q1	Q2	mean	Q3	max	std dev
swipes per user	4	14	21	22.83	28	71	12.08
duration	58ms	236ms	378ms	540ms	638ms	5361ms	522ms

Source: own elaboration

- **2 unique sessions** in different UI configurations, easier, presented in the middle and on right part of the Figure D.1 and second session shown in the Figure D.2. The sessions were differentiated by the number of columns of symbols shown, and a session is differentiated by the ”columnNumber” variable in the dataset.

The basic statistics of the swipes captured are presented in the Table 5.2. For some users only a few swipes were captured, but most of the users had more than 14 samples. Some of the swipes captured were very short or long, but due to the limited size of the dataset, no filtering was performed. This is done due to the fact that in the real financial application it would also be hard to differentiate captured errors from valid data.

5.2.2 Classifying on multiple datasets

With the initial results on Touchalytics dataset providing promising results, further study was conducted on other touchscreen datasets, which could prove that the method can work for multiple users with satisfying accuracy. The aforementioned experiment 1 (Section 5.1) on classifying users was repeated on other available datasets from the literature, described in the Table C. All of the datasets were obtained by the author, with varying licenses. Each of the datasets was studied in depth and only the raw source data was used for the feature extraction preceding the comparison. The following datasets were used in the comparison:

- Tochalytics dataset analyzed before - 41 users,
- Sapienta BioIdent dataset - 71 users,
- Serwadda BTAS dataset - 190 users,
- Own dataset - 88 users.

This means that the results are evaluated on a **sample of 390 users**, where the data was collected in different countries, from users with different age and race (with own dataset es-

pecially suited for making the gender and age distribution more balanced) and on different devices. There are still differences in the task definition between the applications used for data collection and possible differences due to the number of samples per user. However, the goal of the study is to showcase that a certain level of accuracy, burdened with an acceptable level of errors, is possible to be achieved despite this diversity. Use of more datasets would be possible, but some of them were excluded due to the following reasons:

- UMDAA-02 - due to the fact that UMDAA dataset didn't capture user behaviour in a defined environment, but rather observed his/her interaction with a smartphone as a whole - the use of this dataset was out of scope for this experiment. This would result in considerably lower accuracy of the classification and present a scenario that is outside of the scope for this research. As the goal of this dissertation is to design a method which works in a financial application, the environment is strictly defined. We expect to observe interactions dealing with similar, defined tasks e.g., performing a money transfer. Due to that, use of this dataset is not relevant for this scenario.
- Syed JSS18 - this dataset was obtained in a similar way as the previous, but it contains only aggregated features calculated for the swipes, which consisted of only 17 swipe extracted features, user, device and posture information. This is much less than extracted by our method and the lack of source data collected makes the dataset unusable while evaluating the method.

For the Touchalytics dataset and the own dataset session information was included as a variable, as it pinpointed to a different positioning of UI elements and highly influences the outcome. For the Serwadda dataset, both the vertical and portrait sessions were used and merged. Possible strokes with errors (duration > 10 000 ms, length <= 2) were removed, as the authors did not comment on the possible causes of those errors. For the Sapienta dataset the "touchscreen_experience_level" variable available in the dataset was also utilized - to retain the comparability of results.

For all of those datasets grid search procedure was employed, with parameters the same as in the Table 5.1. The CV was 5-fold, in all of the cases the min-max normalized and unnormalized data was tested. In the case of two datasets: Sapienta BioIdent and Serwadda it improved the achieved results. In all of the datasets XGBoost provided the best results, however in the Serwadda dataset the difference was marginal compared to the RF and only improved in case of multi-swipes classification. In all of those datasets 0,8/0,2 split for train and

test was used, but in case of the own dataset two splits are showcased. Due to the smaller number of samples (about 2000), a 0,6/0,4 split was also presented to include more users in the test subset. This is done because for 7 swipes classification we can only include users, who had at least 7 swipes in the test subset (meaning in the 0,8 split, the user would need to have a minimum of 35 actions registered).

The resulting accuracy on Sapienta, Serwadda and Own dataset proved to be similar as in the case of Touchalytics, as seen in the Table 5.3. This proves that the method can achieve satisfying results with regard to the expected accuracy. However, ML (Machine Learning) classifiers have a wide variety of metrics that can be used while evaluating their performance such as precision, recall and the f-score metric, which were also presented.

Additionally, standard approach utilized so far assumed that multiple labels are used to train a classifier. To provide the needed metrics for EER and prove that the method presented can work in a scenario where only two labels are present (0 - legitimate user, 1 - impostor) an extension of an EER calculation metric was used. In this extension, all other user samples (except from the targeted user) were treated as a 1 class. The approach was based on a scikit-learn authors' implementation[2] for plotting the ROC curve for the multiclass problem. Mean EER observed above 2,5% and around 4% for the Touchalytics dataset, on own dataset the value complies with the requirements for the method. The detailed results, including all metrics are presented in the Table 5.4. Unfortunately, due to the size of the Serwadda dataset we were unable to achieve results for the SVM classifier using CV and grid search procedures. SVM was tested without the grid search, but provided significantly worse results and due to the differences in methodology in omitting the grid search procedure, they cannot be included. However, using only one swipe we still can't say prove error rates are better than with the face recognition on all datasets. Fortunately, this approach can be extended to work as a multi-swipe classification scenario. That way we are calculating EER based on the average probability for the N swipes. The results of these tests are described in the Table 5.5.

On each dataset results seem to meet the criteria for EER compared to the mobile face recognition, as is achieved on average 0,068 % EER. The results were consistent but improvements were observed for the addition of accelerometer and minor/major axes of touch on the own dataset, hence partially meeting the consistency evaluation criteria for the method. To further elaborate on this issue, a comparison of feature importance would greatly help in as-

[2] https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html

sessing the extent of the touch profile uniqueness. Moreover, extending the number of swipes to 7 provided results similar to the current fingerprint recognition methods, with an average EER below 0,01% and maximal EER observed on the dataset below 2%. This in turn means that the designed method meets the [A1] criteria in terms of accuracy and EER measurements required in classification scenario. The method was proven to have consistent and accurate results on multiple datasets. What remains to be answered however is the the feasibility of the implementation. We have proven that the method, given enough data and using grid search for hyperparameter optimization can achieve satisfactory results.

Table 5.3. Description of results achieved in a multiclass classification.

Dataset	Swipes	Method	Accuracy	Multi-swipe Classification Accuracy
Touchalytics	20 808	XGBoost	82,36%	93,24% - 3 swipes 98,28% - 5 swipes 99,48% - 7 swipes
Sapienta BioIdent	14 968	Random Forest (normalized)	83,53%	93,27% - 3 swipes 97,86% - 5 swipes 99,07% - 7 swipes
Serwadda	140 075	XGBoost (normalized)	82,08%	91,3% - 3 swipes 97,6% - 5 swipes 98,79% - 7 swipes
Own dataset	2 009	XGBoost	89,78%	(on a sample of 40 users) 98,62% - 3 swipes 99,99% - 5 swipes
Own dataset (0,6 train 0,4 test split)	2 009	XGBoost	86,71%	(on a sample of 60 users) 94,57% - 3 swipes 96,82% - 5 swipes 98,44% - 7 swipes

Source: own elaboration

5.2.3 Comparison of feature importance

To further elaborate on the RG3, a comparison of the feature importance was proposed. Finding similarities between the importances of particular features over multiple datasets may deepen the understanding of the uniqueness contained within the touchscreen behavioural patterns. We know what features differentiate users well in physical fingerprint biometrics

Table 5.4. Comparison of accuracy and error metrics for one swipe scenario.

Dataset	Method	minEER	meanEER	maxEER	Precision (macro/micro)	Recall (micro/-macro)	F-score (micro/-macro)
Touchalytics	SVM	1,5%	5,77%	12,72%	76,74% / 77,21%	74,91% / 77,21%	75,48% / 77,21%
	RF	1,25%	4,02%	10,00%	80,34% / 80,29%	78,24% / 80,29%	78,77% / 80,29%
	XGBoost	0,7%	3,42%	7,14%	81,30% / 82,36%	80,32% / 82,36%	80,58% / 82,36%
Sapienta BioIdent	SVM	0%	4,41%	13,33%	73,83% / 70,34%	69,55% / 70,34%	70,06% / 70,34%
	RF	0%	2,53%	12,5%	84,99% / 83,53%	82,27% /83,53%	82,92% / 83,53%
	XGBoost	0 %	3,03%	10,39%	80,79% / 80,63%	79,64% / 80,63%	79,89% / 80,63%
Serwadda	SVM*	-	-	-	-	-	-
	RF	0%	2,67%	9,68%	82,81% / 82,68%	81,39% / 82,68%	81,66% / 82,68%
	XGBoost	0,01%	2,41%	8,47%	81,58% / 82,08%	80,96% / 82,08%	81,08% / 82,08%
Own dataset	SVM	0%	13,25%	50%	45,45% / 49,25%	45,09% / 49,25%	42,69% / 49,25%
	RF	0%	0,72%	25%	93,39% / 93,25%	89,91% / 93,25%	89,36% / 93,25%
	XGBoost	0%	3,38%	25%	82,08% / 82,62%	76,94% / 82,62%	77,21% / 82,62%

* - Unfortunately there was no possibility of calculating the results for the SVM with the grid search procedure applied due to the computational complexity on a dataset with over 140k samples.

Source: own elaboration

Table 5.5. Description of results achieved in multiple swipes' classification.

Dataset	Method	Multi-swipe mean EER	Multi-swipe max EER
Touchalytics	XGBoost	3 swipes - 0,63% 5 swipes - 0,09% 7 swipes - <0,01%	3 swipes - 2,37% 5 swipes - 1,31% 7 swipes - <0,01%
Sapienta BioIdent	XGBoost	3 swipes - 0,47% 5 swipes - 0,05% 7 swipes - <0,01%	3 swipes - 3,33% 5 swipes - 1,56% 7 swipes - 0,12%
Serwadda	XGBoost	3 swipes - 0,34% 5 swipes - 0,07% 7 swipes - 0,01%	3 swipes - 8,41% 5 swipes - 6,56% 7 swipes - 1,17%
Own dataset	XGBoost	3 swipes - 0,22% 5 swipes - 0,06% 7 swipes - <0,01%	3 swipes - 9,94% 5 swipes - 4,84% 7 swipes - <0,01%

Source: own elaboration

(arches, loops, whorls and smaller minutiae features such as ridge endings), but the studies in the literature have not developed a similar set of traits for touchscreen biometrics.

The task of comparing feature importances of two ML classifiers, even if they are built on the same type of classification method (such as XGBoost) is not an easy task. Due to the different sizes of the datasets and the distribution of samples the output of traditional machine learning feature importance may differ significantly between the datasets. What is worth noting, is that there is no single definition of "feature importance", there are varying metrics that can be used to present how much specific feature contributes to the model. In terms of the used model (XGBoost), according to the documentation, those include:

- 'weight': the number of times a feature is used to split the data across all trees,
- 'gain': the average gain across all splits the feature is used in,
- 'cover': the average coverage across all splits the feature is used in,
- 'total_gain': the total gain across all splits the feature is used in,
- 'total_cover': the total coverage across all splits the feature is used in.

Due to the fact that we are trying to compare the datasets with a different number of observations we cannot use weight and the metrics with "total" prefixes. That leaves gain and cover

values to be available. Gain accounts for the relative contribution of the feature to the model and the average training loss reduction gained when using a feature for splitting. Cover is defined by the number of times a feature is used to split the data across all trees weighted by the number of training data points that go through those splits. Hence using the gain values would make sense in terms of the feature's discriminative power across the dataset.

Using this approach, we were able to identify the rankings of feature importance for the datasets we used. This ranking for the top 25 features is presented in the Table 5.6. What can easily be seen is that:

- Size and pressure data plays the greatest role in differentiating users. In Touchalytics dataset features from this category were ranked 2,4,5,6,7 and 11, for Sapienta it was 4,5,6,7,10,13 and 22; for Serwadda 7,9,14.
- Differentiating orientation (orientation, mid_finger_or) and a direction of a swipe (hor_vert, direction) is important as swipes in different directions have varying values of other features.
- Features connected with velocity, distance and number of measurements (length) are important, velocity connected measurements are ranked 8,9,19,22,23 (Touchalytics); 9,11,15,16,17,18 (Sapienta); 2,3,4,6,8,17,18,21,22,23,24 (Serwadda).
- Positional features (start and endpoint, ranges of X and Y) are ranked: 14,15,16,18,20,21 (Touchalytics); 8,12,14,17,19,21 (Sapienta); 5,10,11,16,20,24 (Serwadda).
- Added detailed size features (rank 1,6,9,10,14,15,18,23) and accelerometer data (2,3,7,8,11,13,19,21,22,24) contribute to the most of the classification importance on Own Dataset. Other important features included size of touch (size_mean), positional features - start and end X position, and velocity group features expressed by duration of the swipe (time) and mean speed.

Comparing the sizes of these datasets, in the task of user identification, we can say that pressure and size measurements are always important in classifying users. Measuring the changes of speed along with the position of swipes influences the classification especially on a large user base (such as Serwadda dataset). Features identified in the own dataset differ due to the fact that X and Y axis separate speed measurements, angle information and pressure data were not extracted for this dataset. Changing the data model to include this information along with additional variables introduced, to further on improve the classifier, remains a valid research goal for further work of the author.

Table 5.6. Comparison of feature importance rankings based on an average gain metric for the XGBoost classifiers.

Feature rank	Touchalytics	Sapienta BioIdent	Serwadda	Own Dataset
1	mid_finger_or	gender_1	hor_vert_v	maxMajor
2	pressure_mean	touchscreen_experience_level	s_iqr	cor_yz
3	orientation	orientation	x_speed_iqr	y_avg
4	size_mean	mid_area	y_speed_iqr	gender_1
5	mid_area	size_mean	start_x	size_mean
6	pressure_med	pressure_mean	a_iqr	major_range
7	mid_press	pressure_range	pressure_mean	x_avg
8	s_mean	xrange	a_q3	dev_avg
9	real_dist	length	size_mean	minor_range
10	direction_4	mid_press	end_x	major_avg
11	pressure_range	x_speed_avg	start_y	cor_xz
12	size_range	start_y	direction_4	start_x
13	direction_atan	pressure_med	direction_atan	z_avg
14	start_x	end_y	pressure_med	maxMinor
15	xrange	vel	direction_3	minor_avg
16	end_y	real_dist	xrange	size_range
17	session_id	time	length	end_x
18	end_x	y_speed_avg	line_dist	minMajor
19	vel	start_x	direction_2	dev_range
20	start_y	yrange	end_y	time
21	yrange	end_x	x_speed_avg	z_range
22	line_dist	pressure_iqr	a_q1	z_iqr
23	length	mean_angle	mean_res_len	major_iqr
24	direction_3	hor_vert_v	yrange	dev_iqr
25	hor_vert_v	largest_deviation	mean_angle	s_mean

Source: own elaboration

5.3 Experiment 3 - authentication scenario

- **Goal:** method's evaluation in authentication and fraud detection scenario. Discussing error rates differences achieved in authentication scenario, as opposed to the classification problem, is inline with Section 4.3.
- **Datasets used:** Touchalytics, Sapienta BioIdent, Serwadda BTAS, Own dataset.
- **Criteria:** security [A1] - in terms of EER achieved in authentication scenario on multiple datasets. Possibility of providing risk metrics for fraud detection as of [B1] requirement and the explainability of the method's prediction required by the privacy [A2] criterion.
- **Description of the experiment:** data from all datasets is preprocessed by the method to obtain features. Three classifiers are evaluated, and achieved results are described by EER and other metrics such as precision, recall, f-score to provide insights into method's performance.

In the previously conducted experiments we have provided answers to a question on the classifier accuracy and consistency when working on multiple datasets. We proved that in accordance with [A1] requirement the method can provide satisfactory accuracy with the error rates lower than mobile face recognition. Secondly, it presented which features differentiate the users and may may be used effectively for building classifiers, proved by the evaluation of relative feature importance comparison across multiple datasets. This provided answers to the RG3 - identifying variables which help the most in creating unique user patterns that may be used for authentication.

Calculations of the classifiers' performance metrics were usable in the identification (classification) scenario and showcased to what extent the model works on a large sample of users. This is due to the fact that if the model can effectively identify a user compared to a number of other users - its characteristics of these users' touchscreen interaction patterns are sufficiently unique. This proved that touchscreen biometrics may be a usable method for biometric authentication, and perform even better when enriched with accelerometer and detailed touch size (touch axes) information. But an adaptation of the model to the authentication scenario requires to give a valid measurement of its performance and observed error rates. The differences between the scenarios have been shown previously in the Figure 4.1.

The previously employed approach to calculation of EER was not exactly fitting for the realistic scenario of a mobile application use. In case we have multiple users (N = 100 for example),

the classifier with an accuracy of 98% on average recognizes 98% of user's samples correctly. If each user had 100 samples, we have 100 valid user samples and 9900 fraud cases for each user in our example. 98% accuracy (assuming equal distribution of errors) for all users would then mean that 0,98 of users samples were correctly predicted as 0 (user) and $1 - \frac{0,02}{N}$ of non user samples were predicted as 1 (anomaly), meaning with multiple users we end up with 2% FRR but FAR is not 2%, but is closer to 0. This in turn makes the calculation of EER biased due to the class imbalance problem. It happens because the classifier tries to maximize accuracy in a multi class classification scenario and to calculate the EER we binarize the classification to the user/impostor problem, and as such we are dealing with the unequal representation of those two classes. The user class (0) accounts only for about 1/N (where N is the number of users) samples. The model also achieved high maximal EER ratio for some users, which is not connected with user's patterns being non-unique, but is rather a result of the optimization strategy of the algorithms. Since the algorithm tried to classify the largest number of samples correctly, it was more prone to fail for the user with a low number of data samples. The min and max EER and the average achieved may be used as a comparison to the approaches described in the literature that were presented before in those scenarios - both in multi-class accuracy metrics and the EER with the inclusion of every other user data. Nonetheless, determining true error rates of the method requires another approach. Summarizing, the previous experiment had a few drawbacks from the perspective of a financial application scenario:

- Identification vs authentication - the method was tested in an identification (classification) scenario, and authentication is more suitable for this case.
- Highly unbalanced classes - the ratio of user to impostor data should be 1:1 or 1:2 to give a valid EER measurement which would be akin to the standard way in which biometric classifiers are evaluated.
- Error rates and decreasing classifier performance - due to the approach being formulated as a classification problem, increasing the number of users increases the complexity of training the classifier and results in a decrease in accuracy. The approach proposed must not decrease its performance due to the number of users.
- Performance and classifier learning - the process of training the classifier needs to be quick to be applied on the device.

Based on those assumptions, an experiment representing an authentication scenario is used. Method tries to learn the user profile (on his/her data) and compare it to an equal (or

slightly larger) portion of another user's data, which represents impostors. In our case we train the classifiers which recognize user (0) or a random impostor sample - an anomaly (1). Using 1:1 ratios instead of 1:N we aim to make the resulting FAR/FRR and EER metrics better in meeting the [A1] criterion. Additionally, this means that for each user the classifier is trained separately, what creates a valid approach for a mobile application authentication procedure. For each of the classifiers user data is be re-labelled as 0 or 1. Due to the performance constraints we use different parametrization for the grid search procedure (shown in Listing 5.2), with a higher learning rate and a lower number of estimators. For this experiment 0.2/0.8 train/test split is used. This means that on own dataset only users with a sufficient number of samples are chosen - for the purpose of classification, their samples may still be used as impostor data. This approach from machine learning perspective regards transformation of the problem from classification to the binary classification problem creating separate models for each class and undersampling the data not belonging to the user class.

```
{'min_child_weight': [2],
'gamma': [0.5],
'subsample': [0.8],
'max_depth': [5, 10, 15],
'learning_rate': [0.1, 0.2, 0.3],
'n_estimators': [100, 200] }
```

Listing 5.2. Grid search parameters for the authentication scenario.

In the scenario presented above we are dealing with only two classes of data, user-0 and an impostor-1. Each of the methods we used can produce the probability of assigning a given sample to a specific class. By the use of predicting probability for each (or a number of) observation we can assign a level of risk to a transaction. In our authentication scenario we evaluated a portion of samples for each user in 1:1 ratio. From the output probabilities we can characterize the performance of the classifier. As can be seen in the Figure E.1 we can observe the errors, the probability calculated for each sample and the resulting ROC curve. Unfortunately for the users with small number of samples the ROC curve may be skewed towards 0.99 or very small values depending on one or two classification results. We can calculate accuracy, precision, f-score etc. from the classification matrix. Using the ROC curve and the resulting probabilities distribution shown in the middle of the figure we can extract the optimal EER.

The resulting performance of the classifiers is presented in the Table 5.7. The method provides below 1% EER for 3 swipes and $\leq$ 0.1% EER for 5. This observed jump in performance is a result of two main reasons: 1) we randomize the distributions of impostors, meaning user is unlikely to be tested on a pattern similar to his/her own. This causes even lower error rates

on a large datasets such as Serwadda. 2) The algorithm learns the user's pattern better due to the better class balance and smaller number of samples. Larger number of users do not influence the classifier's accuracy negatively. This means that in a traditional biometrics testing scenario with 3 or 5 actions the method meets the A1 criterion for performance even without accelerometer and detailed touch major and minor features. Changing the impostor ratio to be bigger e.g., 2:1 compared to user data makes the error rates even lower (see appendices Table E.1).

Table 5.7. Comparison of average method performance and macro averages in an authentication scenario with 1:1 impostor data ratio.

Dataset / Measure	Touchalytics	Sapienta BioIdent	Serwadda	Own Dataset
Accuracy	93,52%	94,34%	94,25%	91,26%
Precision	94,96%	96,06%	97,13%	95,30%
Recall	91,89%	92,51%	91,21%	86,33%
F-Score	93,37%	94,14%	94,03%	90,15%
EER	6,52%	5,94%	5,71%	7,66%
N=3 Accuracy	98,36%	98,78%	98,61%	97,43%
N=3 Precision	98,89%	99,06%	99,03%	98,61%
N=3 Recall	98,85%	99,00%	98,99%	98,23%
N=3 F-score	98,85%	99,00%	98,99%	98,26%
N=3 EER	0,89%	0,72%	0,48%	<0,01%
N=5 Accuracy	99,62%	99,67%	99,56%	99,14%
N=5 Precision	99,74%	99,72%	99,76%	99,54%
N=5 Recall	99,74%	99,69%	99,76%	99,38%
N=5 F-score	99,74%	99,69%	99,76%	99,37%
N=5 EER	0,06%	0,03%	0,02%	<0,01%

N = number of swipes used for classification

Source: own elaboration

5.3.1 Fraud detection possibilities

While the previous sections focused on the performance and error criteria in terms of EER and accuracy, there is still a valid question whether this method can be utilized for fraud detection. To answer it, and directly contribute to meeting the [B1] criterion, a few examples are provided.

The classifiers output probability that a specific observation belongs to a class. Due to that, we can assign this probability to each classified action (swipe), for example:

1. having 3 samples (swipes) labeled as: 0 (user), 0 (user), 1 (impostor),
2. each of the samples passes through the method, which results in a probability estimates: 0.85, 0.55, 0.21,
3. to each of those estimates we can assign risk inversely proportional to the observed probability of a sample belonging to a 0 class (user), or simply proportional to the sample belonging to a class 1 (impostor).

This simple test proves that for single or multiple actions, the risk value could be determined in a way presented in the Equation 5.1, where $P_i(X_1)$ is the probability that the i-th sample belongs to a class 1 (impostor) and N corresponds to the number of samples we are classifying. The resulting probability is the mean probability of the samples belonging to an impostor - therefore a risk of unauthorized access.

$$Risk = \frac{\sum_{i=0}^{i=n} P_i(X_1)}{N} \tag{5.1}$$

However, a question about the explainability of the decision for the customer in accordance with the GDPR still remains. Fortunately, a tool that may be used for this is SHAP (SHapley Additive exPlanations) (Lundberg & Lee, 2017). Is is a new approach for attaining explainability in the machine learning models that was introduced in 2017. It provides game theoretic approach to explain the output of any machine learning model and can also provide predictions on a single row level. SHAP has two distinctive functionalities for our use case:

- It can provide explanations on how the specific variables influenced the prediction, as can be seen in the Figure 5.2. On the right side we see mean importance over the samples used for learning in case of a sample user (id=1) from the Touchalytics dataset that was presented in the Figure E.1. While the left side of the picture provides an aggregated value similar to the feature importance (but is done on test data and not on the model), the right side of the figure presents which variables had high contribution on a specific sample. While the color of the dots presents the value of the variable, placements on the X axis represent the impact on the model input. For example high mean size was helpful in the predictions achieved, same as low mean pressure and high mid area.
- Explaining decisions for a single sample can be seen in the Figure 5.3. The model output was classified as an impostor (which was also the ground truth). The output probability of the sample belonging to a class 0 (user) was about 0,01. The SHAP values explain that

the anomaly was detected mostly due to the low mean size and range of pressure during touch, as the plot was created from the perspective of belonging to 1 (impostor) class. This means that SHAP offers **explainability of the method's decision**.

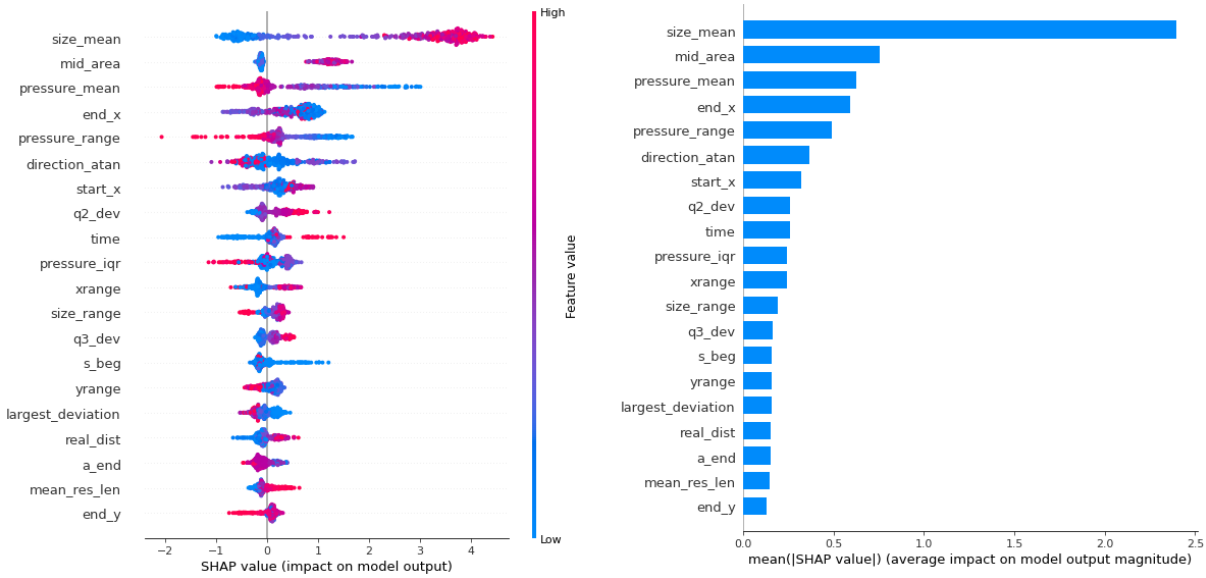


Figure 5.2. Examples of SHAP derived feature importances on all training samples of user 1 from the Touchalytics dataset. Source: own elaboration

Summarizing, by using the output probability we can assign risk values which correspond to the probabilities the underlying ML model in the method produces. By using SHAP it is possible to provide explanations as to why certain decision was made by the method, resulting in achieving high explainability of the model. This in turn means that it is possible to use the designed method for risk assignment on session or transaction level meeting the [B1] criterion and also achieve explainability required by the GDPR and in turn privacy [A2] criterion.

5.4 Experiment 4 - method extension, use of taps

- **Goal:** extension of the method - authentication based on tap actions classification.
- **Datasets used:** own dataset (tap events only - 1497 rows).
- **Criteria:** security [A1] - in terms of achieved EER in the authentication scenario.
- **Description of the experiment:** data is preprocessed and a proposal of features to be included in this scenario is provided in the Table 5.8. XGBoost classifier is employed due to the small dataset size.

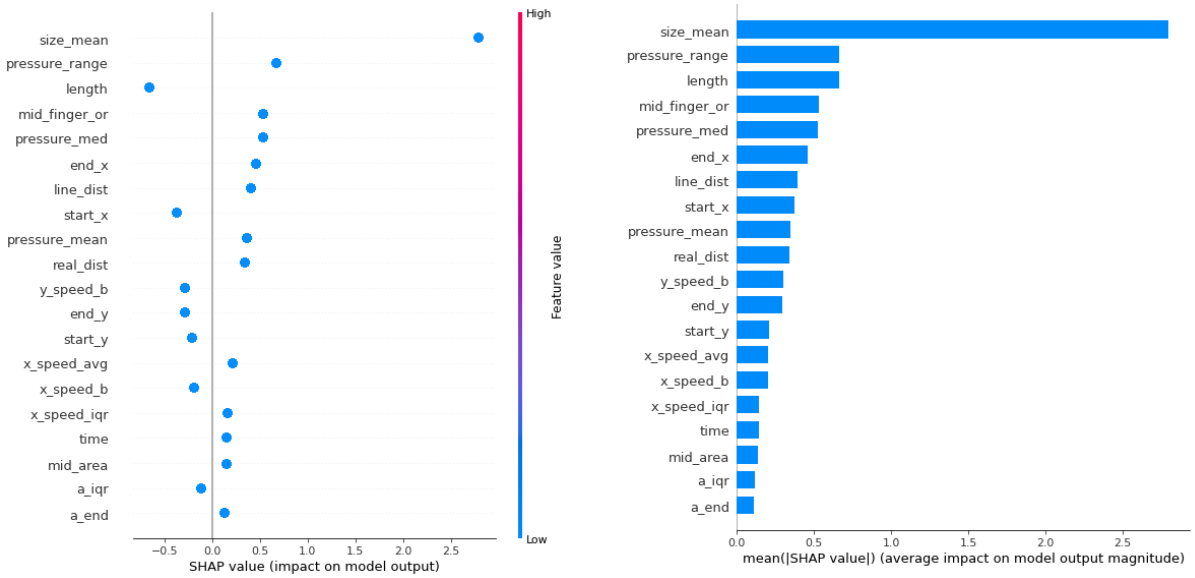


Figure 5.3. Examples of SHAP derived feature importances on one impostor sample against user 1 profile from the Touchalytics dataset. Source: own elaboration

The application designed also assumed collecting information about simplest gestures of users - the taps. Unfortunately, none of the datasets previously used for the experiments contained tap gestures. From other datasets mentioned, BrainRun dataset does contain different gestures, but it lacks crucial touch size and pressure data. Hence, the experiment on using tap data was prepared only using the own dataset. The click dataset contained only 1497 data points for 88 users.

For the own collected dataset, features related to major and minor axes and distance offset to the clicked elements were introduced. All of the features extracted for a single tap included: 'age_group', 'user_id', 'gender', 'columnNumber' (session_id), 'centerX', 'centerY', 'clickX', 'clickY', 'distanceOffset', 'time', 'touchSize', 'yrange', 'xrange', 'size_mean', 'size_iqr', 'size_range', 'maxMajor', 'minMajor', 'maxMinor', 'minMinor', 'major_avg', 'major_iqr', 'major_range', 'minor_avg', 'minor_iqr'. Especially the distance offset variables (distanceOffset, yrange, xrange) are interesting as their introduction relied on results from feature importance analysis of the previous experiments. The columns "deviceModel", "moveType", "age" and "sex" are dropped from the classifier. The features used in this experiment, suited for the extension of the proposed model are described in the Table 5.8.

The distance offset value, along with the explanation of both touch axes is explained in the Figure 5.4. Based on this value, also x axis distance from the center of the touch (xrange) and y axis distance (yrange) could be calculated. As starting and ending positions seemed to

Table 5.8. Features extracted to extend method with tap events.

Feature group	Feature name	Description
Positional	[1]clickX, [2]sclickY	The X and Y coordinates of the click.
Positional (offset)	[3]centerX, [4]centerY, [5]distanceOffset, [6]xrange, [7]yrange	Coordinates of a central point of the element that was clicked [3-4], the calculated Euclidean distance between the click coordinates and the element [5] and corresponding y and x axis distances.
Time and duration	[8] time	Duration of touch event.
Size	[9] size_mean, [10] size_iqr, [11] size_range	Size (based no TouchMajor and TouchMinor elipses) measurement average, median values, IQR and range.
Touch size minor and major ellipse	[12]maxMajor, [13]minMajor, [14]maxMinor, [15]minMinor, [16]major_avg, [17]major_iqr, [18]major_range, [19]minor_avg, [20]minor_iqr, [21]minor_range	Minor and major axis of the ellipse for the touch event in the Android API.

Source: own elaboration

identify users well, the exact placement of a finger could potentially be a useful feature in tap authentication. This approach relies on the idea that users have natural tendencies to click buttons appearing in application with different positioning of their finger.

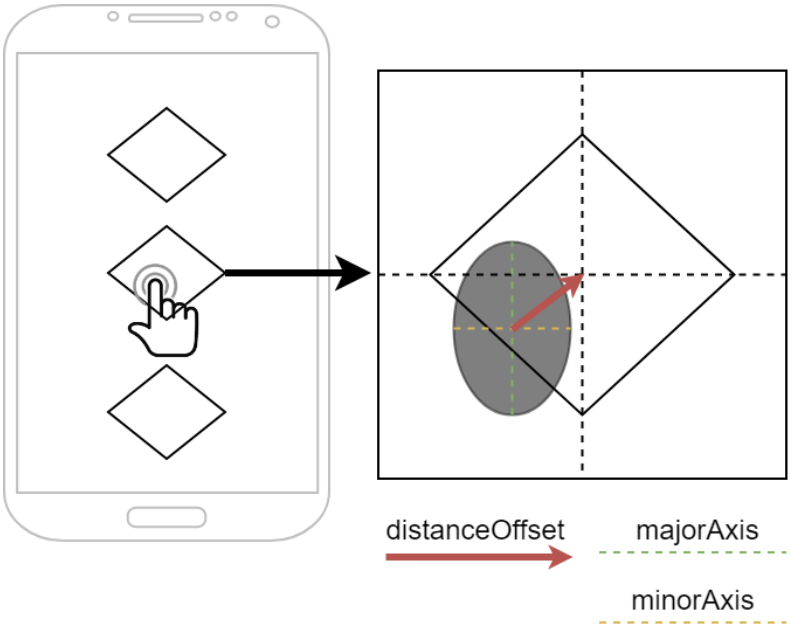


Figure 5.4. Representation of features extracted to address taps, emphasizing the distance offset. Source: own elaboration

The resulting experiment on the own dataset provided 98% accuracy on 0.6/0.4 split with the XGBoost algorithm used for classification. The most important features and the loss function during the learning process can all be seen in the Figure F.1. The most important features identified in this scenario were connected with the size, the minor and major axes of touch and the xrange value which is effectively the X axis part of the distance offset. There was little variation in captured mean size of touch, as can be seen in the Figure F.2, which represents the distribution of mean size variable across the user base. Each column represents one user, as they are sorted by their average size of touch observed. The potential min-max variation captured by the whiskers for each of the users seems rather rare. Users differ by the finger size, which is to be expected for the group that is a representation of different age and gender groups. The achieved accuracy is very high (it accounted for < 0,01% EER), surpassing 3,5% EER for 10 users and 25 attackers on one touch event as achieved by (Inoue & Ogawa, 2018). This highly positive results however are limited by the following reasons:

- Due to the small but varied sample and low number of samples per user, the finger size is the most differentiating factor. There are not many users with the same minor and major axis of touch values, which have not been tested before in the literature.
- These results were observed only on author's dataset and further study is required - either by expanding the dataset to more samples and users or conducting another independent studies.

Summarizing, the use of taps in experiment on the own dataset proved to be a highly successful. Utilizing tap data containing detailed touch size information may improve the classification accuracy and minimize the errors. It also proved that it is possible to use the designed method's approach for tap events with small changes in calculation of features. What could be elaborated more is if the designed method can be used as a proof of presence algorithm and recognize potential authentication insider threat cases, which is the goal of the next experiment.

5.5 Experiment 5 - age and gender recognition

- **Goal:** extension of the method - providing additional information for fraud detection. Recognition of age and gender.
- **Datasets used:** own dataset, BrainRun dataset.
- **Criteria:** fraud detection [B1] - recognizing authentication insider threat, unauthorized access to the device by a spouse, children and people close to the owner of the device and potential to be used as a proof of presence authentication.
- **Description of the experiment:** data is preprocessed and a limited number of features is calculated (BrainRun doesn't contain information for every feature group). Sample of data containing age and gender from BrainRun dataset is used, along with the own dataset for classification of gender and age using a XGBoost classifier.

The method presented showcases an acceptable level of security without exposing confidential user information. Nonetheless, age and gender is often very basic data - given not only to banks, but nearly every mobile application provider. The recognition of user's gender and age group can support the financial application owners in the following areas:

- Recognizing authentication insider threat - unauthorized access to the device by a spouse, children and people close to the owner of the device corresponding to the [B1] fraud detection criterion.
- Detect cases of intentional account/application sharing by multiple users, which may be an interesting information to the app provider and serve as a proof of presence method. Also detecting such situations is important for the performance of the designed artifact, as in one of the assumptions we expect to have only one owner of the device.

Both of those can be the cases of proof of presence. With all of the datasets in the literature, only two datasets contained information about diverse age and gender groups: BrainRun and the own dataset. It is worth to note that BrainRun dataset's age and gender information are non supervised, declarative and not verified. In the BrainRun dataset the features regarding: acceleration, pressure, size, touch minor and majors, finger_orientation could not be extracted. This might have resulted in the lower accuracy of the classifier achieved on this dataset.

BrainRun dataset The BrainRun dataset describes user's behaviour in 5 mobile web games. The dataset itself contains the information about about 2 thousands of different users performing different gestures in those applications. For each user also self-provided gender and age information is available. Some users have purposely chosen not to reveal their gender and set it as 'unknown', hence the dataset effectively contains 3 possible values in this field. The dataset contains the following information about the observed user gestures:

- moveX - the latest screen horizontal coordinate of the recently-moved touch,
- moveY - the latest screen vertical coordinate of the recently-moved touch,
- x0 - the initial screen's horizontal coordinate when the gesture started,
- y0 - the initial screen's vertical coordinate when the gesture started,
- dx - the accumulated horizontal distance of the gesture since the gesture started,
- dy - the accumulated vertical distance of the gesture since the gesture started,
- vx - the current horizontal velocity of the gesture,
- vy - the current vertical velocity of the gesture.

Based on this dataset, from 3 110 102 unique observations representing above mentioned data for various actions, only the user swipes were extracted. This created a sub-dataset that included 1884 users for which user_id could be found (403 female, 1291 male and 190 users with

unknown gender). Overall 646 986 swipes were extracted for those users, which accounted to an average of 343,41 swipes per user. Unfortunately for the use of descriptive statistics the distribution was very long tailed and followed the Pareto rule (20% of the users generated about 80% of the swipes registered), as can be seen in the Figure 5.5. The lowest recorded age was 17.

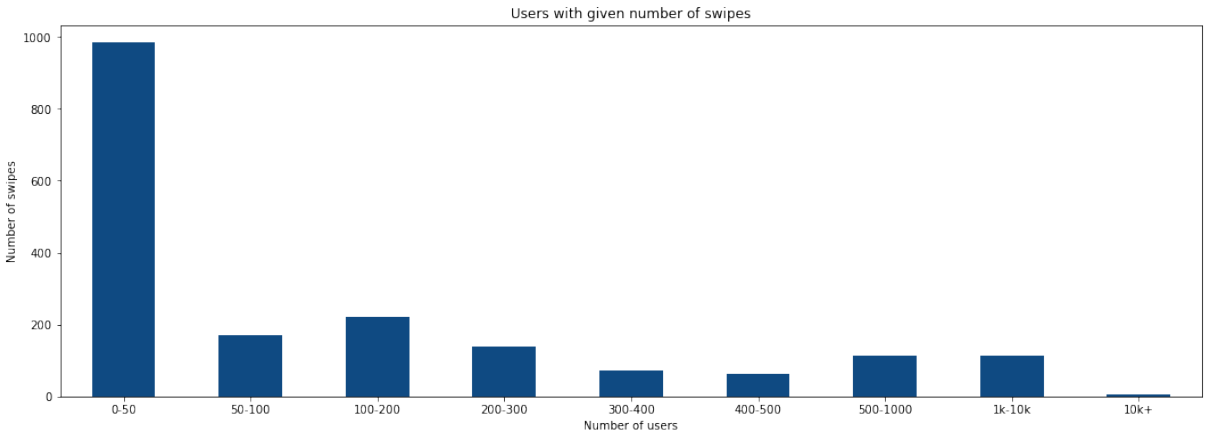


Figure 5.5. Number of users with a given number of swipes collected in the BrainRun dataset. Source: own elaboration

Using the swipes dataset for gender and age classification required extensive cleaning of the dataset. 974 users had no age value, 190 had gender ”unknown”. A few users had age over 100 (which included values like 6565 years) and filtering of such values was performed. However, to prepare a balanced dataset in terms of a number of users in given age groups a filtering approach was needed. Using a sample with e.g., 80% of rows accounting for the given age group would make a classifier try to classify this group properly above all else. Creating equal samples in each age group was impossible - or it would result in a very small number of users in the resulting classifier. As this work's goal was to showcase method accuracy on a reliably big sample which can provide generalizable results - another approach was used.

For each user only up to 400 samples were chosen to use all available data for underrepresented groups like ’60+’ and randomly sample other groups. Next for each combination of age group and gender a max up to 6000 rows were sampled. This gave a representation of 244 users (107 female, 137 male), with the following distribution over age groups:

- 11-20: 25 female, 28 male,
- 21-30: 28 female, 33 male,
- 31-40: 27 female, 29 male,

- 41-50: 16 female, 30 male,
- 51-60: 8 female, 11 male,
- 60+: 3 female, 6 male.

All data accounted for 50 261 swipes, from which 10 053 were used as test dataset (20%). Gender and age classification was then performed on this dataset (referenced later as BrainRun).

5.5.1 Age group classification

For the age recognition, only two datasets in the literature contained diverse age structure of users: own dataset and large-scale collected BrainRun dataset. The age distribution of both datasets in the respective age groups is presented in the Figure 5.6. There still exists an issue of under representativeness of 60 age group and over representativeness of 21-30 group in terms of users, but regarding actions (rows taken) this risk was addressed by the methodology used. The datasets have different sizes, although the data is in both cases extracted from the mobile system API. The device information is not used for the purpose of this comparison of classifier's performance. The resulting XBoost classifier (parameters) resulted in 71,51% accuracy for the BrainRun dataset sample and 88,54% in case of the own dataset. The resulting normalized confusion matrices can be seen in the Figure 5.7. Clearly, under represented age groups seem to fall into 21-30 category, which may only point out to the need for collecting even more balanced datasets. Overall classification accuracy above 70% is not a great result, but nonetheless it was increased significantly on own dataset, which included multiple new features. The resulting accuracy could also be increased in a multi-swipes classification. Due to the fact that the accuracy achieved is similar, the resulting EER would also decrease over the course of 7 swipes. The resulting low accuracy in underrepresented groups can also be seen in the Figure G.1 on the right, where the resulting number of samples is shown instead of normalized classification values. Small number of samples in the last group possibly resulted in low accuracy, compared to the own dataset. The loss function also did not improve after about 200 estimators - which confirmed expected behaviour for a sample of this size.

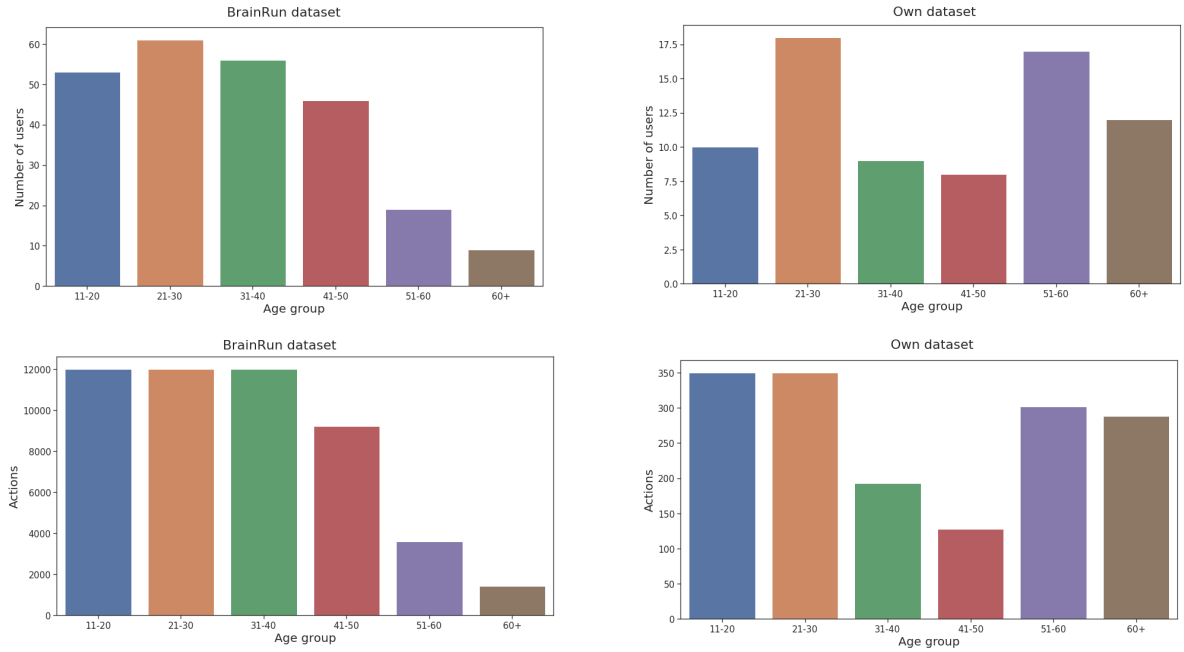


Figure 5.6. Age distribution of users with swipe gestures after filtering for BrainRun (left) and own datasets (right). Source: own elaboration

5.5.2 Gender classification

Similar approach was used to predict gender, with the underlying age_group information given. This resulted in an accuracy of 82.88%[3]. The same approach was carried out for 88 users in the own dataset and resulted in 92,84% accuracy. The results of gender recognition achieved on own dataset, with the use of accelerometer were comparable to the recent study by (A. Jain & Kanhangad, 2019), which achieved similar accuracy values (92% and 93% on their two unpublished datasets).

This increase might have been caused by multiple reasons: better control over the ground truth labels in the set experiment, use of accelerometer and gyroscope data (which was clearly visible in feature importance presented for the datasets in the Table 5.6.) or just a smaller number of users. Nonetheless, the gender classification bounding accuracy of about 80% while utilizing a large scale balanced dataset can be confirmed. It is worth to note that in the BrainRun dataset we could not include multiple features (such as gyroscope or touch axes) which might increase this accuracy in further studies. Especially finger area and minor/major axes could help differentiate gender due to biological issues, which may partially explain differences obtained between those two datasets.

[3]Other ways of sampling the data were used which resulted in the underlying accuracy from 79% to about 87%.

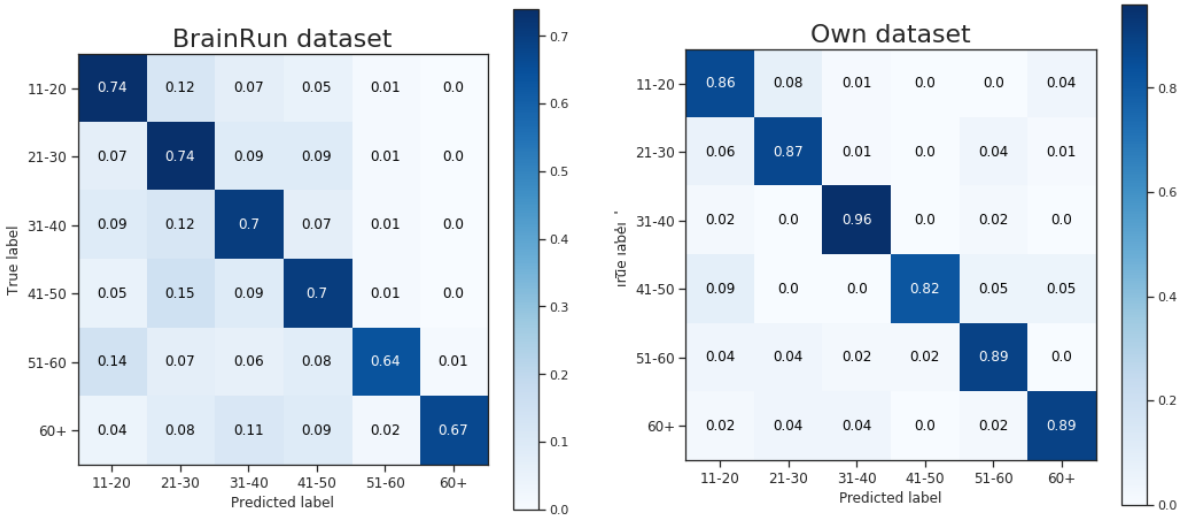


Figure 5.7. Confusion matrix for the age group classification on both datasets. BrainRun dataset on the left, own Dataset on the right. Source: own elaboration

Summarizing outcomes of these experiments, having background knowledge about user's approximate age and gender those classifiers could be trained globally and used to assign additional risk values to the fraud detection systems. This type of risk assessment would not only protect from the cases of theft, but also authentication insider threat. This means that the method may provide additional information to existing fraud detection systems, without exposing user's private information (as gender and birth year are already processed by financial institutions). Coupled with the before mentioned authentication procedure, it may work as a proof of presence authentication mechanism. This, along with the risk assignment on action level mentioned before, means that the methods meets the criterion [B1] for possibility of use in fraud detection systems. It also means that while potential cases of multiple users utilizing the device are out of scope of this dissertation, potential recognition of such cases based on a pre-learned classifier could be possible.

5.6 Evaluation summary

Summarising, all of the experiments were successful, with the only one with sub-par accuracy being age group recognition. The evaluation of the method's performance on multiple datasets allowed us to completely address the research goal **RG3** - by defining and characterizing the importance of the features included in the model. Having in mind which features characterize

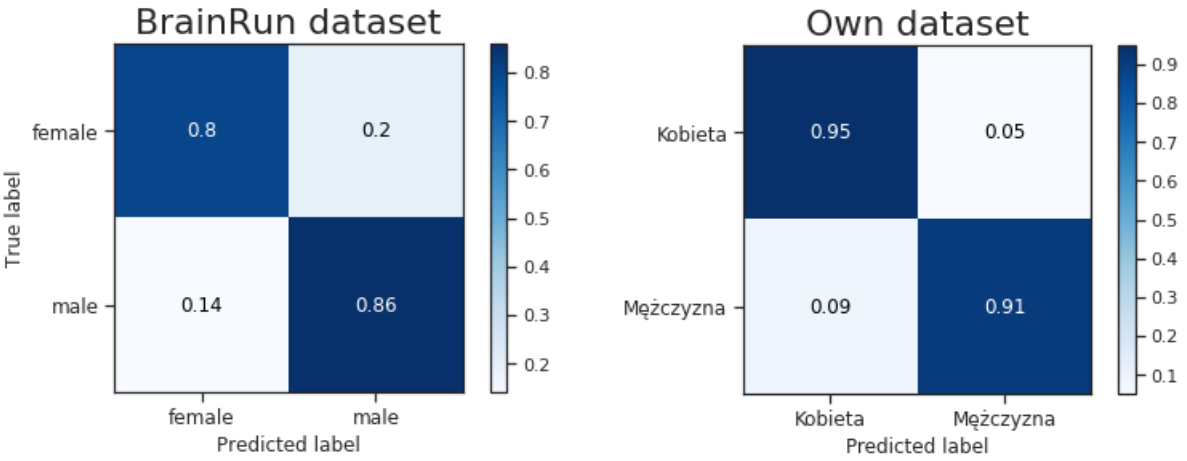


Figure 5.8. Confusion matrix for gender classification on both datasets. BrainRun dataset on the left, own dataset on the right. Source: own elaboration

user's behaviour in multiple datasets, we can verify and characterize the features which may identify and in turn allow to authenticate users.

Having in mind the requirements for the method expressed as RG1, the following criteria have been met by those experiments and the progress of the evaluation is shown in the Table 5.9. The summary of results for each experiment and their influence on achieving the requirements model criteria by the method are described in „Achieved Results" column. The table presents the description of results provided by the experiments in the previous part of the chapter, along with the requirements model they correspond to. In terms of specifying the results achieved by the experiments:

A1 **Security of the method** - based on the defined accuracy and EER criteria for the method, the achieved values were better than those in the literature (listed in the Table 3.7). The results on tap classification proved feasibility of using other actions, but further evaluation is needed on more datasets to provide actionable results in that regard.

B1 **Fraud detection** - based on the method's output conversion to a risk metric, as explained in Section 5.3.1 it has been proven that it is possible to use the method's output in enriching fraud detection systems. The additional possibilities in recognizing age and gender presented in Section 5.5 showcased the method's ability to work as a proof of presence authentication and could be utilized in recognizing insider authentication threats.

B3 **Legal requirements** - the SHAP values presented in Section 5.3.1 allow for the explainability of the method's decision beyond the probability values produced by the classifier's prediction. This in turn means that it is known that e.g., lower speed, change in observed

finger size or other features did not match during the authentication phase and the transaction was labeled as fraud or the access was denied.

Overall, the experiments conducted delivered satisfying results in the context of the above mentioned criteria, although the feasibility of the method's implementation in mobile financial application needs to be studied further.

5.7 Validation - application design

- **Goal:** designing mobile financial application which can use the touchscreen authentication method.
- **Criteria:** usability [C1] - proving that such application can be designed and may benefit from the method's achieved EER and provide better usability during the authentication process.
- **Description of the validation procedure:** analysis of banks' mobile applications design and the number and types of touch actions required to perform transactions. Designing a banking application interface which can benefit from the method's results in terms of actions performed during the money transfer process example.

The dissertation proved that the designed method meets the criteria for performance and the error rates [A1] and can be used for fraud detection [B1] as it provides risk assessment values on transaction level. The feasibility of the method's use in the mobile financial applications remains to be analysed. Potential changes to the current applications are described in the validation part of this dissertation, characterizing the potential use of the method in terms of its interoperability [P1], privacy considerations [A2] and its influence on the usability of the current processes [C1] is the focus of this section.

The goal of this section and the following one, is to provide scenarios which can validate the design of the authentication method that was proposed, beyond the error metrics. First of all, a list of proposed changes for mobile financial applications are provided, which can help in utilizing the developed method for authentication. To propose such changes, an analysis of UI and interaction patterns for the basic functions is be carried out. The design itself, paired with the scenarios depicting the most common functions in those applications serves as a validation for the developed method.

Table 5.9. Requirements and criteria for the method's evaluation - evaluation step.

Banking model requirement, Section 2.3	Method requirement measure	Criteria	Achieved results
Security [A1]	EER (Equal Error Rate) - connected with the FAR (False Acceptance Rate).	$\leq$ 0,08 % EER.	Average 0,05% EER achieved using 5 swipes in the classification scenario in Experiment 2, Section 5.2.2. Average 0,03% EER achieved using 5 swipes in the authentication scenario with 1:1 impostor data ratio proposal in Experiment 3, Section 5.3. About 0,02% EER achieved when using 2:1 impostor data ratio and 5 swipes in accordance with Table E.1. Possible extensions of using tap events presented in Experiment 4, Section 5.4, but since results are achieved on small tap datasets they cannot be as reliable as results with swipes achieved on multiple datasets.
Fraud detection [B1]	Possibility to output probability or similarity metrics on transaction level.	Binary (Yes or No), confirmed with a proof of concept example.	Risk assessment on action level measure presented in Listing 5.1 of Experiment 3. Possible extensions of insider threat recognition by age and gender classification shown in Experiment 5, Section 5.5.
Cost effectiveness and platform independence [B2]	Requiring only system API information and not incurring additional costs on sensors.	Binary (Yes or No), confirmed with proof of concept example that the method can work in a single application and relying on sensors available on nearly every smartphone.	
Legal requirements [B3]	Meeting the PSD 2 requirements for: authentication factor and risk detection. Meeting the GDPR requirements for providing explainability of decisions taken by the model.	Scenario based validation.	Explainability of method's decision shown in Section 5.3.1 of Experiment 3 along with the risk assessment on transaction level.
Usability [C1]	EER connected with the FRR (False Rejection Rate), along with showcasing that the method can be used in financial application.	$\leq$ 1% EER. Scenario based validation.	
Privacy [A2]	Choosing features which are not privacy threatening. Not storing privacy threatening data.	Binary (Yes or No), proved using a data model that does not contain sensitive information.	
Interoperability [P1]	Scenarios of use.	Scenario based validation.	

Source: own elaboration

As an extension of the validation, a comparison of the potential scenarios for data processing is be carried out. Having the goal of describing potential changes in interoperability and privacy characteristics of the method's infrastructure if it was employed in financial environment. Finally, additional scenarios which are unique to the behavioural authentication are provided, presenting the potential of the method to increase usability, provide additional fraud detection possibilities and enable new scenarios of use.

5.7.1 Comparison of chosen banking applications

To showcase the viability of the designed method in a realistic scenario, an analysis of currently used applications is necessary. Our method assumes we should capture multiple swipe motions during the use of an application and completion of the basic functions. Due to that, three mobile bank applications of Millenium Bank, Santander, and PKO Bank Polski (IKO) were chosen for the analysis. More applications were not included both due to the diminishing value and limitations in dissertation size. As the in depth analysis of payment application UI's is not the focus of this dissertation, this topic might be expanded in future research.

The main problem with the adoption of the method to modern mobile financial applications is the interaction process. To provide sufficient level of authentication we need the user to perform multiple touch actions before he/she confirms the transaction or gets access to a function burdened with high privacy risk. The design of modern banking and payment applications, which are the most prevalent type of financial apps is very similar. Each application must have a login screen, as access without confirming identity is not allowed. The login screens for Millenium, Standander and PKO BP applications are presented in the Figure H.1 in the appendices. Choosing a login method is the first part the user must decide when turning on the application. Santander has a separate login only for BLIK services, which is interesting (as those transactions are limited in value), but after a successful login, which is the same as to the main application, user cannot just switch to the main functions of the app. After clicking login button the user is then redirected to the login method, which could be: fingerprint or face biometrics (FaceID), PIN number or a portion of a password (in case of Santander application). This in turn means that those applications work as a point-of-entry authentication and authorization. The login procedure is always present at the beginning of the interaction process.

Further on, after a successful login, the user is redirected to the main menu (see Figure H.2 in the appendices, banks are shown in the same order as previously). In this part of the

application, a user has access to the balance of the account, BLIK services, list of accounts with their history and list of user credit and/or debit cards. The screen is often designed in a way that vertical scroll is needed to access additional functions, although the design may vary slightly depending on the number of accounts, cards and other services. Overall, from this screen the user can gain access to all of the important functions of the application. Since all users land on this screen after the login, it is possible to divide the process of accessing the functions to the login procedure and the path to a specific function. Based on this assumption a quick analysis of touch actions required for them is presented in the Table 5.10. The login process is mostly devoid of swipe interactions, but accessing money transfers requires 2-3 swipes and accessing different products of the user requires at least 1. Accounts balance is always visible, while showing payment history or accessing BLIK service and mostly requires one click. The only difference is if the UI elements require a single scroll to access the button. The money transfer process is very similar between the applications, an example is shown in the Figure H.3 for Millenium bank. Regardless if it is a phone number based or account transfer, user is asked to complete the recipient info, the amount and title. Putting all of the information requires at least one scroll action and clicking the continue button. Finally the user is presented with all details of the transfer, when after a short vertical scroll he/she can finally confirm the payment. Summary of the findings concerning the analysis is presented in the Table 5.10. Main conclusions from this research are as follows:

- While interactions require multiple clicks, only fund transfer and similar transactions (like opening new products) require 2-3 scrolls during the whole process.
- Most of the scrolls are only vertical scrolls, used to access a specific button.
- Access to the BLIK function is available mostly through 1 click after the login procedure.

5.7.2 Touchscreen biometrics continuous authentication

To allow the adoption of the designed method some proposed changes would needed concerning the mobile applications of banks. The proposed idea assumes however that the login procedure (PIN/Password) should not be required for low risk scenarios (for details see Section 5.8.1). The general idea is to utilize different design elements that can be used in the application and contain swipe elements. As such, a proposed design is presented in the Figure 5.9. First of all the login element is replaced by a slider, similar to the one used to answer call. Sec-

Table 5.10. Number of clicks and swipes necessary to access or perform specific functions in the mobile applications.

Bank	Login button	Login procedure	Account balance	BLIK code	Make transfer	Show history / accounts	Other accounts/cards
Santander	1 click	1-N clicks (password) or 1 (biometrics)	-	1 click	4 clicks (+ contact information) / 2 swipes (+ searching for a contact)	2 clicks	2 clicks
Millenium	1 click	4 clicks (PIN) or 1 (biometrics)	-	1 click	3 clicks (+ contact information) / 3 swipes (+ searching for a contact)	1 click / 1 swipe	1 click / 2 swipes
PKO BP (IKO)	1 click / 0	4 clicks (PIN) or 1 (biometrics)	-	1 click	3 click (+ contact information) / 2 swipes (+ searching for a contact)	1 click	1 click / 1 swipe

Source: own development based on the tests of mobile applications on Android platform

ondly, a UI similar to the Millenium bank application is used, but utilizing UI carousel design for presenting credit cards (more graphical version and a use case example for the element can be seen in the Figure I.1). The user would need to scroll down to access money transfer, or use a horizontal swipe. Further on, for the money transfer a popup menu is shown at the bottom, which makes it significantly easier to scroll through the screen for completing the transfer inputs. Further on, both to continue and confirm, the user is asked to use horizontal swipes to confirm the transaction. This design consists of 5 swipes for completing the most basic action of money transfer, one swipe for accessing the account balance or history and at least 2 swipes for accessing other options in the application. Possible another design choices may include a horizontal swipe access menu and swipe up confirmation instead of the horizontal swipe, both of those examples can be seen in the Figure I.2, respectively on the right and left side of the figure. The first option relies on accessing all of the main categories of functions in the application by a horizontal swipe and the second utilizes swipe up for the confirmation of events, which may replace the design presented in step 6 in the Figure 5.9.

What is important, is that regardless of the exact design of those UI elements, the method could potentially not require any login process for accessing these functions, assuming the observed pattern has a risk value (output of the probability metrics for an impostor class) lower than a set threshold. This threshold could be then adjusted accordingly by further tests on a larger scale in a working application. By using the proposed approach, relying on the error rates achieved by the method, assuming about 0,03% of potential frauds (an impostor having access to the user application) we can omit the login procedure 99,97% of the time. This means we

can assume an average of 0,06% EER based on the worst value from the Table 5.7 for 5 swipes and 0,89% for 3 swipes. What is worth to note is that the best achieved EER on own dataset was <0,01%.

Summarizing, after the improvements we assume for the medium and high risk transactions that our process in the designed application takes at least **5 swipes and >2 clicks**, where the number of clicks and swipes may increase if the user transfers funds to a new contact. This in turn, would potentially result in at least **0,06% EER**, based on results from the Table 5.7. The results achieved on own dataset, enriched by accelerometer data proved that this value might be even better and achieved only after 3 swipes. The application designed presented an increase of usability in the scenario of fund transfer due to the continuous authentication during the process. The user does not need to input any credentials and similar design principles can be applied for different application functions. This finally proves that the method can meet the [C1] usability criterion as defined by the requirements model.

5.8 Validation scenarios

With a variety of methods described, a number of possible use cases can be identified, which may apply behavioural biometrics in mobile banking and payment applications. Those include:

- Adaptation of behavioural biometrics methods as a new model of authentication: providing continuous authentication and different authorization levels allowing the standard access systems to be enriched with touchscreen biometrics for high risk transactions - using the *behavioural signature* of a user. This means utilizing continuous authentication and non binary risk output of the method to introduce a risk-based authorization for different actions possible in the application. Use of multiple modalities should follow a general rule that the authentication process should not take more time (or be significantly more demanding) than the task to be performed (Crawford & Renaud, 2014).
- Integration of the methods with current fraud detection systems, which can help in improving the accuracy of current solutions in this field. In this mode the behavioural signature of a user works as an additional factor to a traditional PIN/Password input and may be used to secure high risk transactions, or just increase the security of the transactions and possibly inform the financial institutions of any mis-matches with the pattern to label the transactions accordingly. That way touchscreen profile can be connected with

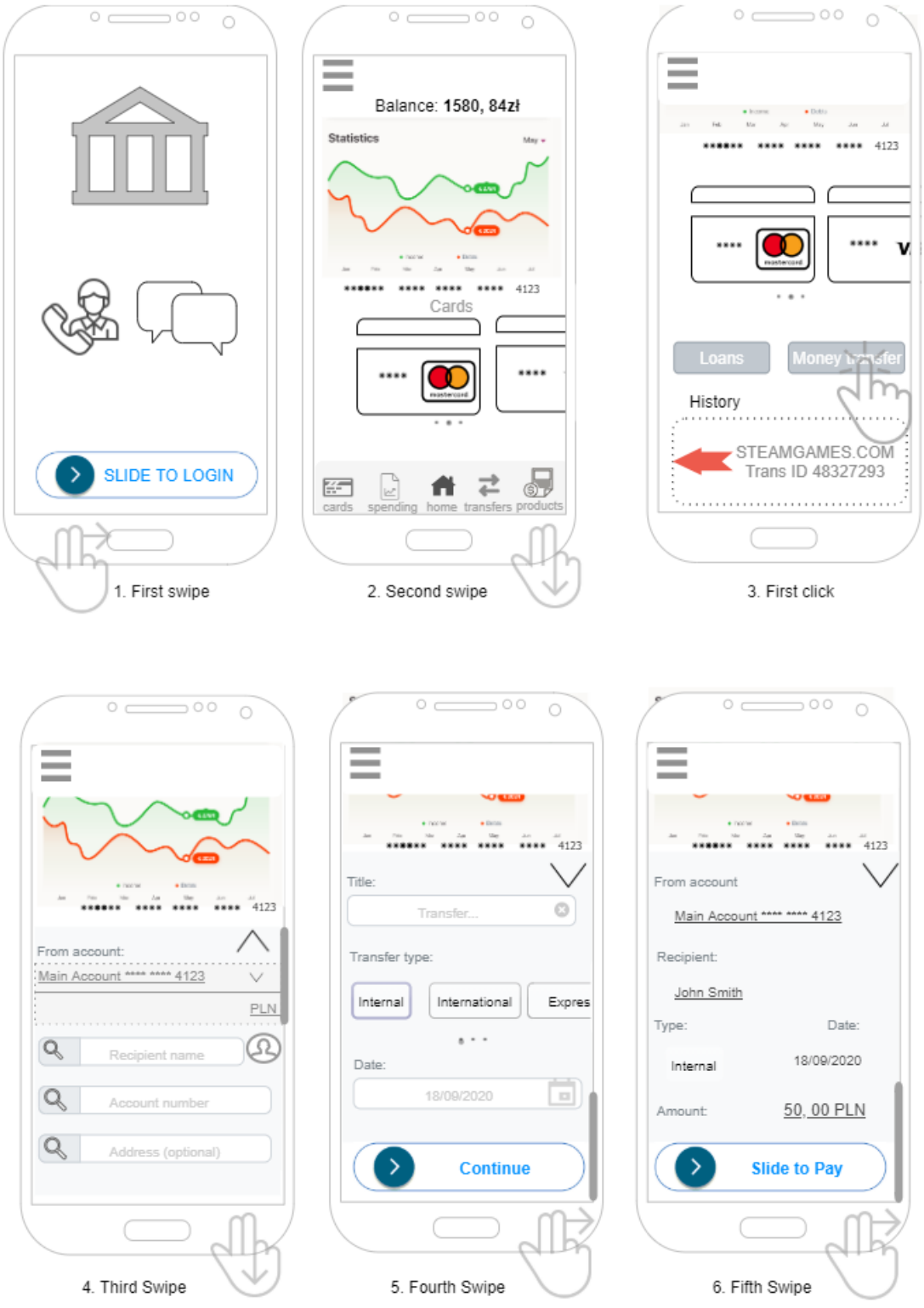


Figure 5.9. Design of a swipe oriented banking app - money transfer process. Source: own elaboration

already known methods to make them ”sensor-enhanced” (Giuffrida et al., 2014). That way accuracy of password or pattern-lock methods can be enhanced with behavioural biometrics signature of a user, which builds up to a multi-modal authentication access system.

Both of these scenarios are important, as they validate different criteria in terms of the method's feasibility of implementation in the financial environment. Due to that, their description will follow a structure similar to the experiments presented previously.

5.8.1 Adaptive authorization and continuous authentication

- **Goal:** utilizing the risk metrics provided by the method to showcase adaptive authentication possibilities with different levels of authorization.
- **Criteria:** usability [C1] - proving that the method can be used in adaptive authentication scenario unlike standard point-of-entry authentication methods. Legal requirements of PSD 2 [B3] - by extending the possibilities of utilizing effective risk assessments on transaction level.
- **Description of the validation procedure:** dividing financial transactions into similar risk groups, proving that the method can provide additional authorization based on the produced risk metrics.

In accordance to the first scenario of operation, due to the non binary nature of the outputted probability, the method can employ authorization on different levels, based on the classifier's output. Meaning, user could be allowed to access some functions without providing additional credentials (like a PIN number), relying on a set threshold of risk (FAR) we can accept. To enable such design, a different risk threshold needs to be assigned to transactions that can be carried out on the device, a simple classification could include:

- Low risk - viewing accounts and cards balance, viewing history of recent transactions.
- Medium risk - defined payments (bills etc.), small funds transfers, transfers to trusted contacts, card actions (block, deactivate).
- High risk - large funds transfers, contacting customer service, ordering new services/-cards, loans.

According to the PSD 2, strong authentication principle requiring multiple factors may be bypassed for low sum or low risk of fraud transfers, cyclical transactions and transactions from

white listed shops and institutions and safe corporate transactions (Europen Comission, 2015). This means that those transactions could be authenticated using only behavioural biometrics - representing a single factor. The risk value would be assigned according to the 5.1 from Section 5.3.1. High risk actions may require additional authentication mechanisms. Customer service contact is included in this category, as for IT systems people seem to be the weakest link of their security. When accessing such actions other credentials should be provided - password, PIN or another layer of authentication. Having in mind this division of actions on low/medium/high risk, a proposed simple design for an adaptive authorization is presented in the Figure 5.10. Different probability levels that translate to the confidence the samples processed in the method belong to the 0 class (user) - inverse of risk, are written above the three different designs. Firstly, all functions, including the account number are available, user only needed to swipe to get past the login screen. Secondly, due to the slight mismatch, when the algorithm is not sure a PIN number could be required either to gain access or complete certain functions (like money transfer, applying for loans etc.). Some medium risk transactions could still be possible without further authentication, having in mind that risk value changes continuously when the user performs more actions in the application. Finally, when the pattern does not seem to match, the application may start - but access to any of the functions is possible only if the user provides a PIN code. All of the confidential information (full card or account number) should be hidden or partially hidden. The algorithm assumes that if a user inputs additional credentials (PIN/password/fingerprint) the application sets its level of confidence to the one corresponding to 100% of options available. Nonetheless, the situation when the a pattern was not recognized could then be transferred to the bank institution fraud system to asses the potential problems with the method itself.

5.8.2 Fraud detection integration - data processing scenarios

- **Goal:** utilizing the risk metrics provided by the method to showcase fraud detection possibilities and evaluating the possible implementation scenarios.
- **Criteria:** interoperability [P1] - by evaluating the design of two different architecture proposals for the integration of the method by external applications and describing the feasibility of implementation. Fraud detection [B1], privacy [A2] and partially cost effectiveness and platform independence [B2] - by designing scenarios of communication

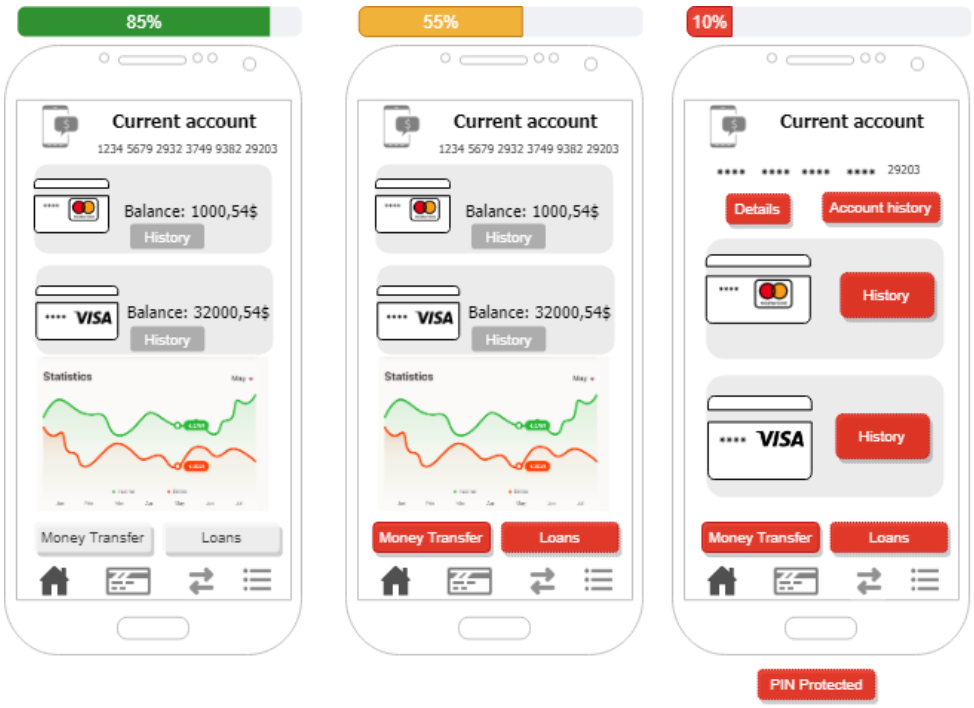


Figure 5.10. Example of varying authorization levels calculated based on a current risk value. Source: own elaboration

with the fraud detection system for 3rd parties along with the privacy preserving data model that minimizes the risk of sensitive data leak.

- **Description of the validation procedure:** designing the privacy preserving data model based on the SHAP output, which can be tied to user identifier when communicating with bank's architecture. Analysis of the method's learning and authentication phase in the financial application. Evaluation of architecture choices for either central or dispersed (edge) data processing in terms of privacy and other requirement model's criteria.

Describing the developed artifact - the authentication method relying on a touch profile of a user, its potential deployment in the financial architecture remains to be discussed. The proposed solution should meet the criterion for interoperability [P1], and in turn should also allow being integrated with existing banking systems and and comply with the Open Banking standards.

Due to the fact that the method utilizes phone data, which is directly accessed from the open system API and available for all mobile applications, it can be processed directly on user's mobile phone. The general design of the method is to first, learn the user profile and only when enough data is collected that it is stable - switch to touch authentication. This change should be as seamless as possible and provide continuous authentication. Due to that mode

of operation, we need to divide scenarios of use for the application in two parts: learning and authentication. Due to the machine learning nature of the model, which can easily be retrained based on available data it should also be able to work in a semi-supervised manner to improve its performance over time, enabling the evolution of the method over time. First mode of operation relies on a set of impostor data against the user provided actions. As shown in the Figure 5.11, first the method would only collect swipes and perform the model training - possibly only in the first initialization round. The labelling of correct actions would be confirmed by PIN (or physical biometric) authentication. Further on the algorithm would continuously compare captured samples with the model and produce predictions about their class. These predictions would then be evaluated based on an EER, like previously mentioned in this work. If the model would produce satisfying results of evaluation, the training process would finish and the model would switch to the authentication mode, presented in the Figure 5.12. PIN authentication would only be required in cases where the method would predict impostor class and require additional authentication factor. After the identity confirmation, the samples could be then transferred to the model again and retrain it. This in turn allows for the constant retraining and evolution of the model in a semi-supervised manner. **The bank or OFI could define maximal risk value (probability of an impostor performing the transaction) for different sets of actions, utilizing adaptive authorization** based on the logic described in the previous section. That would mean different evaluation results would trigger the denial of access, relying on the risk calculated.

The proposed scenarios can assume that false negatives (denies of access for the user) would be detected and flagged properly and false negatives (fraud cases) would then be evaluated on the side of the financial institution. The output probability information along with the SHAP description could be assigned to a fraud detection system with every transaction as the last step. The data model that can be used for presenting this information with all transactions is shown in the source code in the Listing 5.3. The presented data format includes basic information about the transaction and user application, the anonymized SHAP values assigned to specific features e.g., ”F1”, and prediction results - risk levels (0-1 range), connected with each of the swipes. The final risk value is then calculated based on multiple swipes just like in the multiple actions classification presented in the previous chapter.

```
{”TransactionID”: ”########”,
”ApplicationID”: ”########”,
”Timestamp”:     ”########”,
”Swipes”: [
```

```
  {"Timestamp": "########", "Screen": "########",
  "Risk": 0.04
  "SHAP_values": {"F1": 0.32, "F3":-0.25, "F23": 1.32},

  {"Timestamp": "########", "Screen": "########",
  "Risk": 0.01
  "SHAP_values": {"F6": 0.42, "F3":-0.15, "F23": 0.98},
  ...
],
"final_risk": 0.02
}
```

Listing 5.3. Basic data model for edge processing communication with fraud detection.

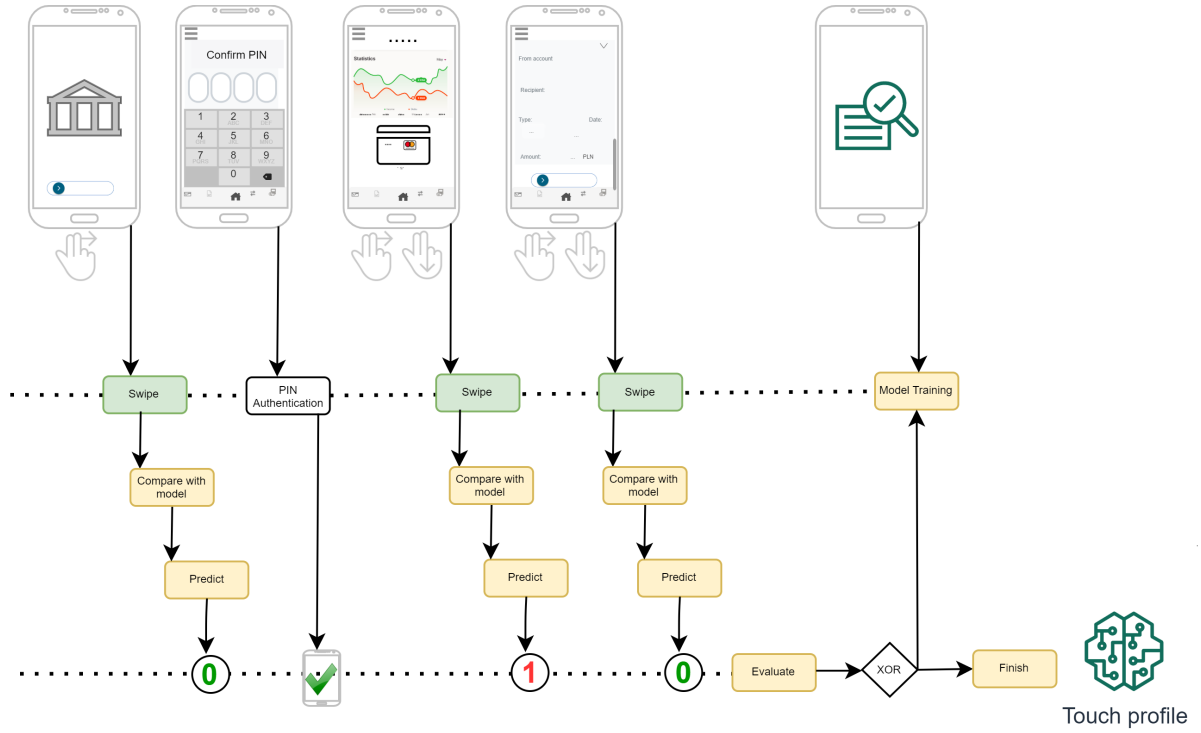


Figure 5.11. Method implementation in an authentication scenario - learning phase. Source: own elaboration

Communication with the fraud detection system after a function is accessed (e.g., funds transfer is completed), as presented in the Figure 5.12, is preferred due to the privacy reasons. That way, the data is transformed locally and the classifier is trained in the same way, minimizing the risk of privacy threat and pattern leak. This showcases the local - on the edge computing implementation model, where the only information processed is shown in the proposed data model.

On the other hand, this is not the only mode of operation for the proposed method. Due to the fact that the data processed is characterized by a low privacy threat due to processing only device captured events inside the application, it is possible to also process the authentication and authorization process on the provider's infrastructure. Mobile application requires

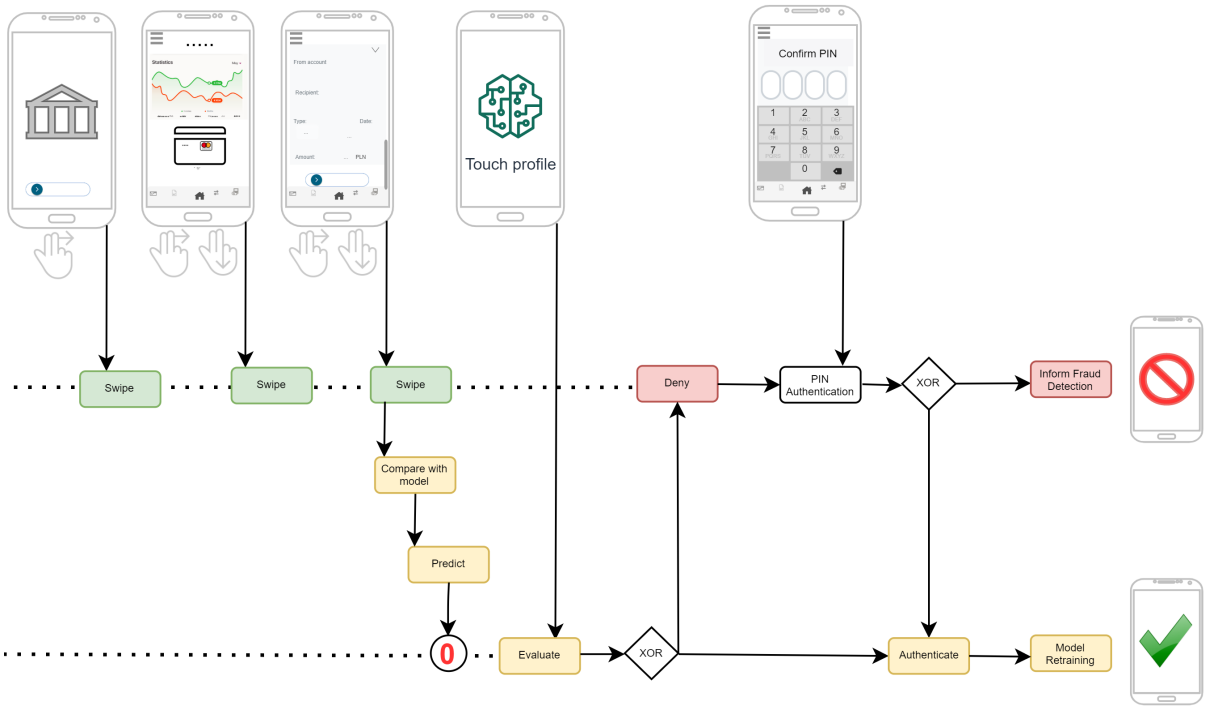


Figure 5.12. Method implementation in an authentication scenario - authentication and re-training phase. Source: own elaboration

constant access to the Internet and must send encrypted data to the financial infrastructure nonetheless. This means that after the device extracts the features from the raw data and characterizes each swipe (or possibly other actions) the data is then sent to the central processing node on the financial provider's architecture. Both of these scenarios are different and valid ways of deploying the solution in terms of the interoperability [P1] criterion. As they are connected with different advantages and drawbacks in terms of privacy [A2] and cost effectiveness [B2] criteria, they need to be at least briefly analyzed.

Those two scenarios of data processing can be tied with different levels of private data exposure and the responsibility of data storage. The communication scenarios when requesting authentication are also different for both these implementation proposals and are presented in the Figure 5.13. As such, in the first case only the fraud detection system is outside of the application, meaning the risk of exposing any confidential information is tied to the device security. In the second case also the database (called TouchDB) is stored on the provider's side meaning that storing the data becomes an issue and may in turn be prone to data leaks. This however means that the service provider is more confident that the touch data have not been tampered with. Both of these architectures have advantages and disadvantages, which have been presented in the Table 5.11. The overall process begins with the swipe itself, where

the application collects the data about the touch event from the Android API. Further on the variables defined in the Table 4.2 are extracted by the application controller. This, along with the prediction and risk assessment can either be done on the device in first scenario of on the edge computing, and on external infrastructure in case of the central processing architecture. The data about the risk is either saved in step 8, after the process is finished or directly to the fraud detection system infrastructure in step 7 for the central processing scenario. The biggest issue of the on edge processing concerns providing a good representation of imposter data, either generatively or through internal tests supplied with every version of the method and sent to the application model. On the other hand, central processing stores the data that can be used for replay attacks and increases the threat of a pattern leak if more institutions were to utilize this infrastructure. But what is important, if the infrastructure would only store anonymized features with the use of encryption and some kind of user level salt, the possibility of pattern leak is still minimal and the re usability of the pattern is also not burdened with high risk.

Table 5.11. Comparison of the architectures for the method's implementation.

Characteristic	On the edge computing	Central processing
Privacy [A2]	Higher, data can be stored on the device only temporarily	Lower, risk of potential data leak considering stored profile and data sent
Interoperability [P1]	Higher, due to the fact that the method can be tied to a single library installed on the application	Lower due to the need for communicating with financial institution architecture
Cost effectiveness [B2]	Low for maintenance and use, High for learning and method improvements	Higher for maintenance and use, low for learning and method improvements
Performance	Lower, with additional problems of supplying impostor data for the classifier	Higher, better accuracy and training possibilities for the model
Main advantages	Very high security and little to no privacy risk	Decision explainability and potential improvement of the model over time is significantly easier due to the data storage
Main drawbacks	Problems with updating the model, algorithm changes and method improvements	Higher costs for infrastructure maintenance, higher risk of not properly securing the communication channel, higher privacy risk in terms of pattern leak

Source: own elaboration

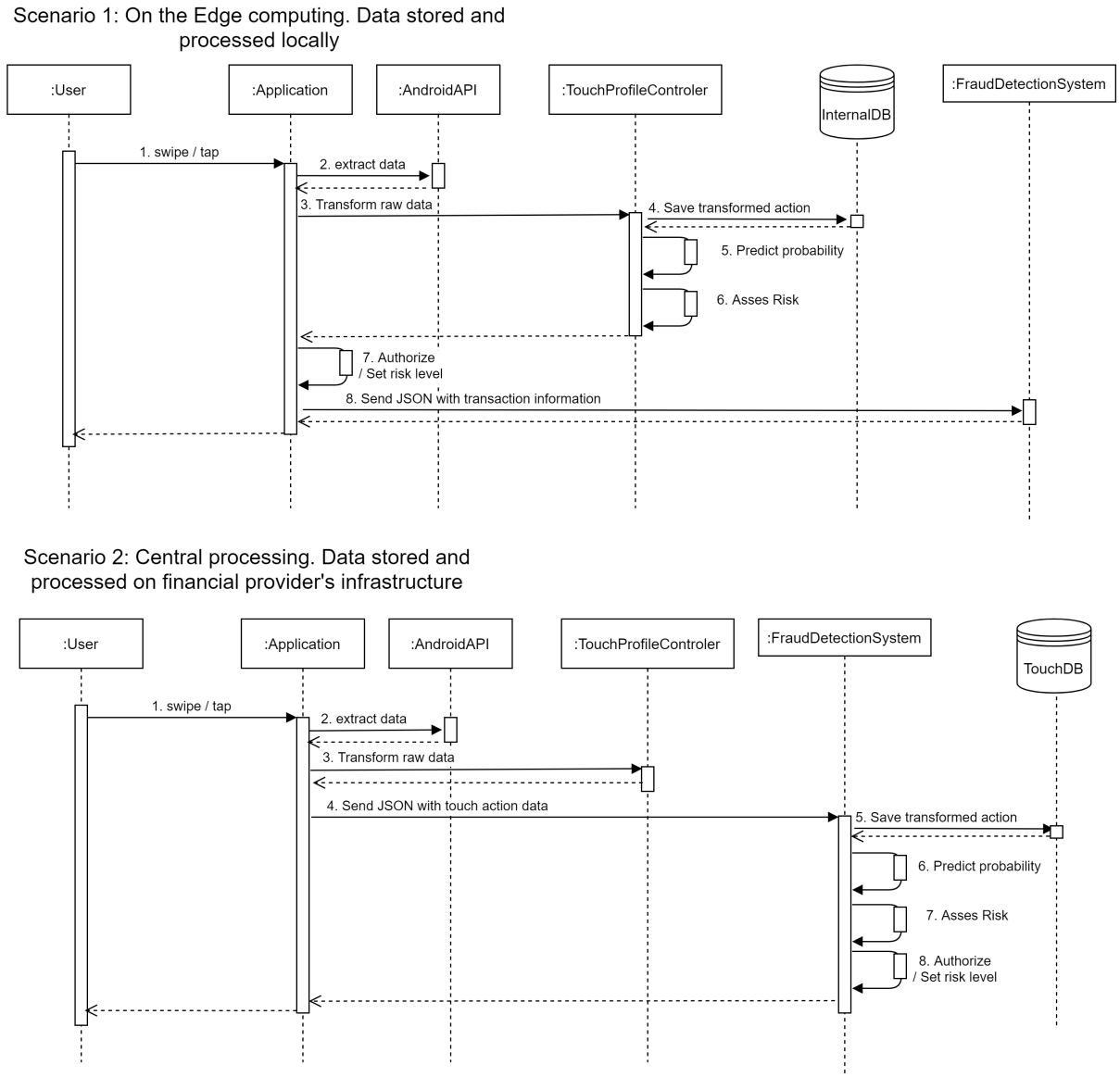


Figure 5.13. Data processing scenarios during the authentication and retraining phase. Source: own elaboration

5.9 Summary

The research presented in this chapter showcases that it is possible to effectively authenticate users of mobile applications with the use of own touchscreen based method. The achieved error rates meet the security [A1] criterion if multiple actions are sequentially classified. The approach was described in depth in the Section 5.6. The method achieves accuracy better than the studies in the literature (including the most recent studies such as (Hernández-Álvarez et al., 2021)) and also better than mobile face-detection methods. By the addition of custom features, including the accelerometer readings, the method adds to the state of the art.

Additionally, the method proposed discussed the use of taps with measurements of touch size. The achieved result is very promising and also meets the performance criteria for the method. Further validation is however still required for the use of taps as no dataset in the literature employed similarly extensive number of provided features. With the use of swipes it was also possible to classify the gender and the age group of a user, although with only about 80% accuracy in the second case. Nonetheless, with gender classification providing very promising results, evaluated on a large datasets, it is possible to differentiate in a between multiple users of a phone. Age classification in age groups has shown that it can be performed based solely on interaction patterns in the application. As the method employed can provide probability estimate that a user belongs to a given age group or gender - this information can be then passed on to the anti-fraud system. Same can be of course applied to the user authentication score. Proven possibility of enriching systems with this information meets the criteria for the method aimed at fraud detection and risk assessment. The gender and age recognition experiments also prove that method can be applied in the scenario of biometrics insider threat detection and possibly as a proof of presence method.

Further on, the validation scenarios for the method were proposed, which answer directly to the RG5, proving the feasibility of implementation in the mobile financial environment. Having in mind the requirements for the method, the following criteria have been proven by those experiments:

C1 - Usability: based on the average 0,03% EER achieved by the method previously for 5 swipes, its feasibility of integration needed to be tested for a mobile financial application. By providing the application design in Section 5.7 an improvement of usability when using financial application was presented and proven. It was then extended to actions which

require fewer actions by providing possibility of assigning different authorization levels as described in Section 5.8.1.

B1 - Fraud detection: based on the risk output and the possibilities of explaining the classification by SHAP algorithm the method can asses risk on transaction level. The scenarios for the method's communication with the fraud detection system were showcased in Section 5.8.2 with two architectural choices with own advantages and drawbacks.

A2 - Privacy: the proposed privacy preserving data model was presented in the Listing 5.3 that allows method implementation with little to no risk of exposing user data. On the other hand, central processing scenario was presented to pose low level of threat, due to processing only touchscreen interaction data, which could be further encrypted.

P1 - Interoperability: the presented scenarios along with the data formats prove the feasibility of the method's implementation in an Open Banking scenario. The data format presented can be exchanged by the API infrastructure and the method is possible to be employed as a single library utilized by every third party mobile application requiring touchscreen authentication.

The overall possibility of implementation including the scenarios in which the method could train the classifier and evaluate the samples - providing continuous authentication and adaptive authorization (extending the possibilities in which the [B3] criterion for PSD 2 risk assessment was met) was shown in the Figures 5.11 and 5.12.

Summarizing, the presented results of evaluation and validation are compliant with the model presented in RG1 as presented in the Table 5.12. The results achieved by the validation scenarios and not showcased previously in the Section 5.6 have been written in bold. These results in turn **fulfill the RG 5**, with accordance to its contents, **evaluating the method's performance and validating it in scenarios applicable for mobile financial applications**. The method's feasibility of implementation was proven in the chosen environment by the designs of the architecture presented. Additionally, the evaluation of the method's variables performance allowed us to completely answer the research goal **RG3** - by defining and characterizing the importance of the features included in the model across multiple datasets, as shown in the Section 5.2.3.

Table 5.12. Requirements and criteria for method's evaluation - validation step.

Banking model requirement, Section 2.3	Method requirement measure	Criteria	Achieved results
Security [A1]	EER (Equal Error Rate) - connected with the FAR (False Acceptance Rate).	$\leq$ 0,08 % EER.	Average 0,05% EER achieved using 5 swipes in classification scenario in Experiment 2, Section 5.2.2. Average 0,03% EER achieved using 5 swipes in authentication scenario with 1:1 impostor data ratio proposal in Experiment 3, Section 5.3. About 0,02% EER achieved when using 2:1 impostor data ratio and 5 swipes in accordance with Table E.1. Possible extensions of using tap events presented in Experiment 4, Section 5.4, but since the results are achieved on small tap datasets they cannot be as reliable as results with swipes achieved on multiple datasets.
Fraud detection [B1]	Possibility to output probability or similarity metrics on transaction level.	Binary (Yes or No), confirmed with proof of concept example.	Risk assessment on action level measure presented in the Listing 5.1 of Experiment 3. Possible extensions of insider threat recognition by age and gender classification shown in Experiment 5, Section 5.5. **Scenarios for communication with fraud detection system for 3rd parties and proposed data model described in the Section 5.8.2.**
Cost effectiveness and platform independence [B2]	Requiring only system API information and not incurring additional costs on sensors.	Binary (Yes or No), confirmed with proof of concept example that the method can work in a single application and relying on sensors available on nearly every smartphone.	**The proposed application design showcased in the Section 5.7 presented that the method can be applied relying solely on touch sensors API. Different architecture proposals for deployment described in the Section 5.8.2.**
Legal requirements [B3]	Meeting the PSD 2 requirements for: authentication factor and risk detection. Meeting the GDPR requirements for providing explainability of decisions taken by the model.	Scenario based validation.	Explainability of method's decision shown in the Section 5.3.1, Experiment 3 along with the risk assessment on transaction level. **Expanded risk assessment possibilities in adaptive authorization scenario included in the Section 5.3.**
Usability [C1]	EER connected with the FRR (False Rejection Rate), along with showcasing that the method can be used in a financial application.	$\leq$ 1% EER. Scenario based validation.	**Achieved maximal EER 0,06% of 5 touch events and $\leq$ for 3, validated to be achievable in a mobile financial application and provides increase in usability. Detailed description was provided in the Section 5.7.2. Extension of usability improvement over transactions requiring lower number of actions was presented in the scenario of adaptive authorization in the Section 5.8.1.**
Privacy [A2]	Choosing features which are not privacy threatening. Not storing privacy threatening data.	Binary (Yes or No), proved using a data model that does not contain sensitive information.	**Privacy preserving data model presented in the Section 5.8.2, along with the analysis of architecture proposals in terms of sensitive data processing.**
Interoperability [P1]	Scenarios of use.	Scenario based validation.	**Deployment scenarios with two different architecture proposals were presented in the Section 5.8.2.**

Source: own elaboration

Chapter 6

Summary and outlook

6.1 Research results and contribution

The challenge presented in this dissertation was to characterize the requirements and build a new method of authentication that can improve the security and usability of current mobile banking and payment application authentication process. The method should help in solving the most important issues of the sector, including the issue of rising CNP frauds, adoption to the mobile-centric model of banking and security requirements to provide new services in this model.

Inline with that, the main research problem of the dissertation was that financial services require authentication methods suited for mobile application environment, which could enrich current fraud-detection systems on transaction level. They also require methods, which could offer security not hindering the usability of the process itself compared to the currently used methods. As the nowadays employed method of mobile face detection offers only point-of-entry authentication, its usefulness in terms of offered usability can be significantly improved by other methods.

The method should, according to the thesis proposal, be integrated into mobile banking and payment scenarios and provide error rates comparable or better to the currently utilized mobile face detection methods, while increasing the usability of the solution. The work presented in the dissertation achieved this result by meeting the research goals and answering the research questions, namely:

RG1: The requirements for the authentication method successful implementation in the financial sector were presented in Chapter 2 in Section 2.3.4.

RG2: The suitable method was chosen to be touchscreen behavioural biometrics. This choice was based on listing the possible sensors that can be used in behavioural authentication methods and the broad description of results. Relying on the descriptions of characteristics of the chosen methods in Chapter 3, summarized in the Section 3.5, touchscreen behavioural biometrics were used as best suited to the requirements model.

RG3: By listing the potential features used in Chapter 3, it was possible to design the main artifact of this work, the authentication method relying on touchscreen biometrics. The features used in the method's design were listed in the Table 4.2. The contribution of specific features was tested on multiple datasets in Chapter 5, Section 5.2.3. This in turn allowed for identifying and choosing the features which offer the best possibilities to characterize the user, regardless of the scenario and application design.

RG4: Main artifact of this dissertation was the authentication method, which was designed in the Chapter 4. The developed method was tested in multiple experiments and was proven to achieve below 0,08% EER in the authentication scenario for 5 swipes used in the process, as presented in the Table 5.7 in Chapter 5. The use of accelerometer and touch axes in the method allowed achieving similar error rates for only 3 swipes, which was shown on author's dataset.

RG5: The scenarios used for benchmarking of the methods were presented in Chapter 5, comparing the method in multi class classification (identification) and binary classification (authentication) scenarios. This showcased the potential differences in EER that can be achieved by the design of the method scenario. Based on that, the validation was performed by designing the scenarios in which the method can be applied in the financial application, described in Sections 5.7 and 5.8. Potential architecture designs described the learning and classification process, possibilities of method's evolutionary capabilities by retraining in semi-supervised scenario and described data processing scenarios in which the method might be employed. Presented communication schemes and deployment proposals also proved the feasibility of the method's implementation in financial environment.

Detailing the requirements model of RG1, as its importance for the method's design was crucial, method met all of the categories mentioned:

- Security [A1] - the error rates achieved by the method were proven to be less than 0,08% EER, according to the results of authentication scenario from the Table 5.7, with the as-

sumption of 5 swipes before the process takes place. The method may be used to prevent device theft, transactions unauthorized by the user and potentially also malware masquerading as the phone owner.

- Usability [C1] - the method was proven to increase usability in more than 99% of situations according to the EER achieved, employing the approach presented in the Figure 5.12 assuming functions which require 5 actions before confirmation in the mobile application. The potential of improving the usability is also possible to be employed from the first action performed by the user, by using the approach presented in Section 5.8.1. The method provides continuous authentication and by assigning different risk levels by adaptive authorization.
- Privacy [A2] - relying on the features used in the method, described in Chapter 4 no sensitive information is used in the authentication process. Relying on the edge processing scenario possibility, risk of exposing user pattern would be minimal. Even in central processing scenario, the only thing possibly exposed would be used touch information and probability estimates of the method with feature importances - which can change with the device itself and might be encrypted on the provider's infrastructure. This in turn limits the problems of the pattern revocability and permanence.
- Legal requirements [B3] - the designed method enriched with accelerometer readings is classified as a biometric inference factor by the literature in Chapter 3 and the review on banking methods in Chapter 2. European Banking Authority also considers behavioural biometrics as a valid inherence factor in PSD 2 required authentication procedures (EBA, 2019). This means the method is considered valid for authentication in financial environment and meets the criteria for begin a factor of strong customer authentication (SCA) in accordance with PSD 2 directive. It can also be safely employed in risk assignment of singular transactions, allowing for the potential of employing the proof of presence authentication.
- Fraud detection [B1] - the method can provide probability estimates to the user's identity and the possibilities to assign risk scores associated with single transactions, as was presented in Section 5.3.1. It also allows for the explainability of the decisions due to the use of SHAP, which is required to provide compliance with the GDPR. The scenarios in which it might be employed were showcased in sections 5.8.1 and 5.8.2. The method also

analysed the possibilities of employing insider threat prevention mechanisms by proof of presence authentication, which was proved in Section 5.5.

- Interoperability [P1] - the method enables the estimation of risk probability for individual transactions on the basis of assessing the uncertainty as to the user's identity when using the application. With little risk of private information disclosure, FinTech companies could use this method to communicate with banking infrastructure in a transaction risk assessment scenario. The developed method does not rely on any particular hardware or principles which might be hard to employ. The proposed data format and the infrastructure in which the method could work in were presented in Section 5.8.2.
- Cost effectiveness and platform independence [B2] - the method requires no additional sensors and relies on the most basic operating system APIs. Meaning it can be used on nearly every touchscreen ready device. The tests in this work covered multiple different models of phones from year 2011 to 2020 and proved to achieve very similar results.

Summarizing, the method allowed for meeting the thesis assumptions that: *an authentication method designed using behavioural biometrics can be deployed in a mobile financial application and achieve error rate lower than currently employed mobile face detection methods, providing higher usability.*

The results achieved by this dissertation from the perspective of the SotA on behavioural biometrics have contributed to the knowledge base in a few different areas. First of all, this work designed a method for financial environment, developing a technical financial innovation, which can be used for authentication, risk assessment of transactions and can prevent a portion of CNP fraud cases in the mobile environment. This method can offer continuous authentication and risk based authentication allowing for different risk levels .

Secondly, the work presented showcased the differences achieved by the researchers in terms of comparing multiple touchscreen behavioural biometrics methods and defined a research gap, which was only partially answered by meeting the thesis assumptions. The design of the evaluation process for the authentication methods in mobile biometrics, considering the length of the learning process, design of the data collecting application and the time required for authentication, allowed to develop a set of experiments which compared the performance of the method on multiple datasets. The results achieved showcase that there is a possibility of achieving very low error rates for behavioural biometrics with the use of novel classifier such

as XGBoost, but multiple measurements are needed. As such this work contributed in terms of analysing the research methodologies and experiment designs of the different researchers, and proposed an unified evaluation environment considering the use of a set number of actions. It also showcased the differences that may be achieved in multi and binary classification scenarios, with the same method and datasets.

The designed artifact provides a mobile touchscreen biometrics method which employs an extended array of features, as described by the Section 4.6, which can be used by other researchers. The most novel of the proposed variables are connected with the accelerometer and touch axes. The importance of the features presented was compared among the datasets, what, to the best of the author's knowledge, has not been performed in the literature before. The use of accelerometer and potential employment of taps for classification was also mentioned tested on own collected dataset. The results may still require confirmation on multiple datasets, due to the limited number of measurements and users available on own dataset, but the achieved accuracy was significantly higher than for any other dataset. The work also discussed the possibility of utilizing gyroscope readings to improve the classification, and provided the respective features, but was not employed due to the datasets limitations.

6.2 Limitations of the research and further work proposal

Currently the method deals with the problem of user authentication and risk assignment based on user touchscreen behaviour. The results achieved are satisfactory and meet the required criteria for the use in financial sector. However, to achieve them the method relied on some assumptions, which were required due to the time and complexity constraints of the dissertation. First of all, only one user was utilizing the phone and no sharing is present during the learning phase. This assumption was presented in Chapter 4. Recognition of such situations could be performed using one-class classifiers and anomaly detection techniques during the learning phase. Automatic recognition of multiple profiles is an interesting area of further research and may be important in terms of insurance compensation for frauds. As of now we require the user to use his or her phone alone we could employ creation of separate profiles/accounts for multiple family members, but it would lengthen the learning process. Recognizing multiple user could also be done by the usage of one class classification approach and anomaly detection algorithms. Due to the fact that the algorithm assumes we have impostor data sup-

plied, assumption of only user data present (one-class) could allow for improving the ease of implementation of the method. There are examples in the literature which showcased similar accuracy with one-class classifier methods as early as 2006 (Mazhelis, 2007) and practical data-driven approaches to dealing with this issue are present in the recent literature on behavioural biometrics and anomaly detection (Choi et al., 2018; Kumar et al., 2018; Yang et al., 2019). The introduction of one-class classifier would solve most of the problems observed in the on the edge computing scenario. For now the migration between devices for now would require relearning the user profile, as the profile is built to be device-bound. Each time the device is changed application would need to enter the learning phase as described in 5.11. This is a desirable trait based on the privacy requirement of GDPR and potential risk of pattern leak. We have not studied minimal number of actions which were required to build a stable profile, but in most cases utilized data from about 5 sessions (Touchalytics and Own dataset) and average of 25 actions during a session was enough to achieve satisfying results. In this environment the algorithms have proved to be effective in user authentication, so the learning process may take only a few days. Further work on this topic is however important and should be carried out after acquiring additional data with all of the described touchscreen variables available. Further studies in the topic could also include translating the points to the relative measurements to avoid changes caused by the multiple devices observed.

Secondly, the variability of the observed profile in a long term and its stability was not studied. By a proxy of mobility profiles studies some methods could be adapted to this area. Unfortunately for now most of the datasets do not cover lengthy time periods and research in this topic would require more data collection. This however could be great area of further study, although requiring significant funding to collect datasets which cover a year or more of a user's activity. This work also did not test different body positions of a user while utilizing mobile phone as showcased by the recent study of Syed (Syed et al., 2019). Inter session's stability of the pattern could be studied by extending the collected dataset.

The results achieved on own dataset in terms of taps and observed increase in accuracy can be somewhat attributed to the smaller size of the dataset and not enough session variability presence. While this does not invalidate the results achieved on multiple different datasets, extension of research in this topic could help in validating the importance of accelerometer and touch size axes in achieving low error rates. Also further researching the uniqueness of accelerometer patterns captured inside the applications as an alternative to gait recognition

could offer an interesting research experiment, in which SherLock and BrainRun datasets could be used. The results could not only improve the proposed method, but potentially allow to introduce more haptic oriented behavioural biometrics to the available methods.

Due to the changes encouraged in the UI of the applications which adopt touchscreen authentication, measuring the usability change in an interface prototype can be a valid choice of an applied science approach to extend the study findings. The performance should be measured based on the HCI metrics for touchscreen interaction, possibly in an approach similar to the study by Goguey et al. in 2018 (Goguey et al., 2018). Design of a banking application function was also not performed. Using other gestures and connecting them with possible UI elements - pinching, zooming etc. could also be the topic that could extend the method to more sophisticated cases. Studying the stability of the pattern and the effects of the context on the pattern registration and performance including posture, device size and configuration should be studied, in accordance with the related work (Syed et al., 2019).

As for the possible extensions to the dataset collected, the influence of the different resolution of the device on a pattern captured in the application could be studied. Similarly, collecting more tap data along with the accelerometer and gyroscope readings may help to showcase that taps can be incorporated into the design to achieve even lower error rates that currently presented.

The limitations of this research however do not diminish the achieved results, but offer interesting areas of further research which can expand the results and prove method's feasibility in multitude of different environments and scenarios which were not the focus of the dissertation.

Appendix A

Glossary of terms

- Adaptive authorization - compared to the point of entry authorization it allows giving different levels of privileges based on different authentication factors and procedures which compare valid pattern to the one captured for an authenticated individual (Ayed, 2014). This means authorizing an individual whose pattern doesn't match perfectly may result in access to only a limited number of access privileges to the individual e.g., for low risk actions where we can provide very high usability and accept higher error rates.
- Authentication factor – any object or collection of one or multiple extracted features/characteristics that can be used for performing the authentication process. Often also referred to as the category of credentials used to verify the identity. It may be divided into: knowledge, inherence (biometrics) and possession factors (Bolle et al., 2013).
- Behavioural biometrics – covers biometric features that can be considered unique (or sufficiently distinguishable), non transferable, hard to forget or lose, difficult to reproduce and hide but derive from user behavior rather than physical traits of an individual (Saevanee et al., 2012b). It relies on an observation of behaviour quantifying the way some action is carried out and extend over time (Bolle et al., 2013). It may cover *"Any readable and processable representation of user behaviour which exhibits identifiable and repeatable patterns that can be used for identification or authentication."*. The biometric itself can be based either on a single type of feature expressing the behaviour or a combination of multiple behavioural characteristics/modalities (Mecke et al., 2018).
- Behavioural profiling - group of methods for behavioural biometrics which try to characterize the unique pattern of usage for different services of the device. Focused on

capturing the interactions with the GUI and the readings of the devices sensors ”as is” and not labelling them to any physical behaviour.

- Biometric - is a characteristic of the human body (physical) and human actions and behaviour (behavioural) that can be used to differentiate people from each other (Bolle et al., 2013). A single biometric can be created from a captured pattern of a single feature or a combination of features/traits/characteristics of an individual (different names are used in the literature). Most often when we refer to this characteristic it is tied with a single modality such as the fingerprint.
- Biometrics or biometrics methods - biometrics is the science of identifying, or verifying the identity of a person based on the physiological or behavioural characteristics (Bolle et al., 2013). Biometrics methods utilize unique or distinguishable enough characteristics to perform identification and authentication of people (Saeed, 2012). These processes are done based on physiological and behavioural appearances and the the characteristics that allow to distinguish one person from the next. The biometric methods aim to capture these characteristics - creating distinctive and identifiable patterns, which are the captured and quantified representation of the features.
- Continuous authentication (also implicit authentication) - is the way of continuously processing the user pattern and performing the authentication process of assessing the validity of user claimed identity. Due to the fact that it requires processing the credentials or the extracted characteristics it is often required that to ensure usability the process needs to be transparent and require little to no user interaction.
- Digital banking - is the notion of the digitization of the banking process including every program and activity undertaken by financial institutions and their customers.
- Electronic banking customer - is the user that utilized web, mobile or wearable technology to access banking services.
- FinTech - is the acronym for financial technology. It is used to describe new companies which offer services that seeks to improve and automate the delivery and use of financial services. FinTechs include a variety of companies, which are not banks and large financial institutions which may offer services to the banks themselves (algorithms, methods, new services) or work as intermediates for customers offering them innovative services often relying on a communication with bank infrastructures in Open Banking standard.

Companies from this category are regarded as third parties for the purposed of the requirements model created in Chapter 2.

- Insider threat (authentication) - is a threat of unauthorized access of an individual which is known to the original possessor of the account/device (Hayashi et al., 2012; Muslukhov et al., 2013). This may mean family member, child or spouse. It is similar to the notion of informed attacker from information systems security but may also mean that the attacker posses user credentials and private information which may be used in an attack. One of the ways of preventing insider threat in this sense is the use of proof of presence in the authentication procedure.
- Mobile banking customer - is the user that utilizes mobile or wearable technology to access banking services.
- Mobile-centric banking - is the notion proposed by Deloitte (Srinivas et al., 2018) similar to the digital banking but focusing on the transformation of traditional branch focused banking model to a mobile environment. It assumes all services should be accessed in the banking environment from a mobile device.
- Multi-factor authentication - is the use of more than one authentication factor for the process, where the features belong to at least two types from: knowledge, inherence and possession.
- Multi-modal authentication - is the use of more than one biometric feature/characteristic for authentication, for example: gait + keystroke or voice + face. The features do not need to represent different factors.
- OFI (Other Financial institution) - OFIs include all financial intermediaries that are not central banks, banks, public financial institutions, insurance corporations, pension funds or financial auxiliaries. In this work they are regarded as third parties for identifying the stakeholders in the financial adoption model. They include mainly investment funds, captive financial institutions and money lenders, central counterparties, broker-dealers, finance companies, trust companies and structured finance vehicles but also include FinTech companies.
- Open Banking (standard) - the idea of Open Banking standard was introduced by the PSD 2 directive. Its goal is to incentivise and standardize the access to electronic banking services and data stored by the financial institutions. It ensures that financial institutions create mechanisms to enable third party providers (including FinTechs and other OFIs) to

work securely, reliably and rapidly with the bank's services and data on behalf and with the consent of their customers. Whilst the EU does not explicitly require banks to use Application Programming Interfaces (APIs).

- Physical biometrics - covers features connected with observable physical traits of individuals, which are unique, universal among the populace, measurable - allowing to extract patterns and sufficiently permanent for the use in authentication and identification of individuals captured at some point in time. It includes: fingerprint, hand geometry, retina/iris biometrics, face and vascular biometrics.
- Point-of-entry authentication/authorization - is a standard type of authentication used in mobile devices and most of internet services nowadays. It is a type of authentication and authorisation where if a user provides a credential or is authenticated based on any factor full extent of his/her authorized privileges, for example if use provides a password we are 100% sure about his/her identity and give all the account privileges. Due to the fact that the authentication process automatically assigns the highest available access privileges to the individual, thus concluding the authorization process, the authentication and authorization are used interchangeably in naming the approach.
- Proof of presence - is a rather new type of security proof (Samet et al., 2019) especially used when considering electronic devices. It confirms that the individual that is the identity holder is physically present in the authentication and authorization process during the time it takes place. It can also be understood as a persuasive evidence that any one particular user, rather than another, is directly responsible for carrying out a particular action in the application. Is is important as it allows to discover the cases of using stolen credentials (password, PIN numbers) to access the device and potential cases of insider threat and unauthorized access carried out by for example family members.
- PSD 2 - is the Revised Payment Services EU Directive (PSD2, Directive (EU) 2015/2366), administered by the European Commission (Directorate General Internal Market) that aimed at regulating payment services and payment service providers throughout the European Union (EU) and European Economic Area (EEA) (Europen Comission, 2015). It regulates which type of organizations can provide payment services and to what rules they must comply with. Out of the extensive list of regulations, most important from the perspective of this work are the inclusion of FinTechs as electronic money institu-

tions (EMI), opening of the banks architectures as Open Banking and the requirements for customer authentication and risk assessment of financial transactions.

- Strong Customer Authentication (SCA) - is a concept introduced in the PSD 2 directive that relies on utilizing at least two different allowed methods representing different authentication factors to confirm the authentication of an individual or authorization of an action (despite the name, as the directive states which actions should require strong authentication). The requirement ensures that payments are performed with multi-factor authentication and has been a legal requirement for electronic payments and credit cards since 14 September 2019. Whether strong authentication is mentioned it means SCA. In line with this notion weak authentication does not require two factors.

Appendix B

Biometric adoption in Polish mobile banking applications

Table B.1. Adoption of physical biometrics in the Polish mobile banking applications.

Biometrics in the Polish mobile banking applications.	
Alior Bank	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application since Q4 of 2018. Voice biometrics is planned.
BGŻ BNP	Biometrics used as an electronic signature in one of the channels. For corporate client logging with a fingerprint in iOS and Android (Touch ID), face biometrics used in iOS application (Face ID) since Q2 of 2019.
Santander SA	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application. Available for all customers since Q2 2019. Biometric voice system is used for customer identification. Carries out research regarding video/camera identity confirmation.
Citi Handlowy	Both fingerprint and Face identification. Has own Face identification technology which is used for credit card registration process since Q4 2019.
Eurobank	Acquisition by Millenium in Q4 2019.
Getin Bank	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application. Since Q4 2019 uses ”digital fingerprint” (effectively behavioural profiling).
ING Bank Śląski	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application and custom solution for Android applications. Uses behavioural profiling since Q1 2020 for fraud detection.
mBank	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application and custom solution for Android applications. First to use behavioural profiling in Q4 2019.
Millennium	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application. Has own Face identification technology which can be used account registration process.
Pekao	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application since Q1 2018 (earliest adopter).
PKO BP	Logging with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application. Piloted their own project (Czyżewski et al., 2019) for multi-modal biometric authentication in 2018 (signature, fingervein, voice and face biometrics).
Plus Bank	Logging with a fingerprint in iOS and Android.
Raiffeisen Polbank	Logging and transaction authorization with a fingerprint in iOS and Android (Touch ID) and face biometrics (Face ID) in their iOS application.

Source: own development based on the information available on the banks' websites

Appendix C

Publicly available behavioural biometrics datasets

Table C.1. Publicly available datasets which can be used for behavioural profiling.

Name	Year	Data collected	Volume of data
MIT Reality Mining and Social Evolution datasets[1]	2008	Behavioural profiling - Phone and CDR data + bluetooth, proximity and WiFi. Additional survey data.	Complete data records for 80 users and survey data.
Nodobo dataset[2]	2010	Behavioural profiling - Phone and CDR data + bluetooth, proximity and WiFi data.	Data records for 27 high school students, 5,5 billion rows. users
LiveLab dataset[3]	2010	Behavioural profiling - Phone application usage, CDR, phone mode, accelerometer sensor, wifi data, hashed web browsing history.	Data records from 25 users (college students) using iPhone 3GS observed during the period of about a year.

[1] http://realitycommons.media.mit.edu/
[2] http://nodobo.com/release.html
[3] http://livelab.recg.rice.edu/traces.html

University of Maryland Active Authentication-02 (UMDAA-02) Dataset[4]	2015 (published in 2018)	Behavioural profiling + touchscreen + other sensors (front-facing camera, touchscreen, gyroscope, accelerometer, magnetometer, light sensor, GPS, Bluetooth, WiFi, proximity sensor, temperature sensor and pressure sensor. The data collection application also stored the timing of screen lock and unlock events, start and end time stamps of calls, currently running foreground application).	Complete sensor signals collected from 48 volunteers on Nexus 5 phones over a period of 2 months, 141.14 GB of smartphone data.
Sherlock dataset[5]	2015-2017 (published in 2016 and updated dataset in 2018)	Behavioural profiling -Phone and CDR data + bluetooth, proximity and WiFi, resource utilization per running App (CPU, memory), anonymized Location, Network statistics, malicious software examples.	10 billion data records, massive time-series dataset spanning nearly every single kind of software and hardware sensor that can be sampled from a Samsung Galaxy S5 smartphone, without root privileges.

Source: own development

[4] https://umdaa02.github.io/
[5] http://bigdata.ise.bgu.ac.il/sherlock/#/

Table C.2. Analysis of the mobile touchscreen datasets.

Name	Event types	Aggregation	Position (x,y)	Pressure	Area	Direction	Additional sensors	users / events	sessions for user	devices
Touchalytics[6] (Frank et al., 2012)	scrolling and strokes	-	yes	yes	yes	yes	-	41 users 21k	7 (2 after a week)	4
BioIdent Sapienta[7] (Antal et al., 2014)	strokes	- (raw data available)	yes	yes	yes	yes	-	71 users 14k	4	8
H-Mog[8] (Sitová et al., 2016)	scrolling, strokes, tapping, multi-touch events	-	yes	yes	yes	yes	keystroke, accelerometer, gyroscope, magnetometer	100 users >1.5 million	24 (3 tasks)	10
UMDAA-02 Touch[9] (Mahbub et al., 2016)	scrolling and strokes	-	yes	yes	no	yes	-	48 users >3,5 million	varying (avg. about 200)	9
BrainRun dataset[10] (Papamichail et al., 2019)	tapping and strokes	-	yes	no	no	yes	accelerometer, gyroscope, magnetometer	2218 users >3 million gestures	varying (avg. about 50)	2418 (90% is Android)
Syed JSS18 (Syed et al., 2019)	strokes	On stroke level, features extracted	for start and end stroke	yes (aggregated)	no	yes	-	31 users 600k	≥ 8	4
BTAS 2013 Serwadda (TouchDB benchmark version) (Serwadda et al., 2013)	strokes	-	-	yes	yes	yes	-	190 users > 2 million	2	1

[6] http://www.mariofrank.net/touchalytics/
[7] https://ms.sapientia.ro/~manyi/bioident.html
[8] http://www.cs.wm.edu/~qyang/hmog.html
[9] https://umdaa02.github.io/
[10] https://zenodo.org/record/2598135#.YGmJIegzaUk

Table C.2. Analysis of the mobile touchscreen datasets.

Name	Event types	Aggregation	Position (x,y)	Pressure	Area	Direction	Additional sensors	users / events	sessions for user	devices
Own dataset	tapping and strokes	-	for start and end stroke	-	yes	partial	accelerometer, gyroscope, touch axes	88 users > 3500	2	4

Source: own development

Table C.3. List of the available mobile keystroke datasets.

Name	Year	Data collected	Volume
MEU-Mobile KSD[11]	2016	Keystroke	2856 records, 51 records per user for 56 users
RHU KeyStroke Mobile-based Benchmark for Keystroke Dynamics Dataset[12]	2014	Keystroke	956 records, 17 avg records per user for 51 users
Sapienta MOBIKEY Keystroke Dynamics Password Database[13]	2014	Keystroke	2142 records 51 avg records per user for 42 users
Coakley mobile[14]	2016	Keystroke and Gestures	965 records for 21 users

Source: own development

Appendix D

Design of the application for data collection

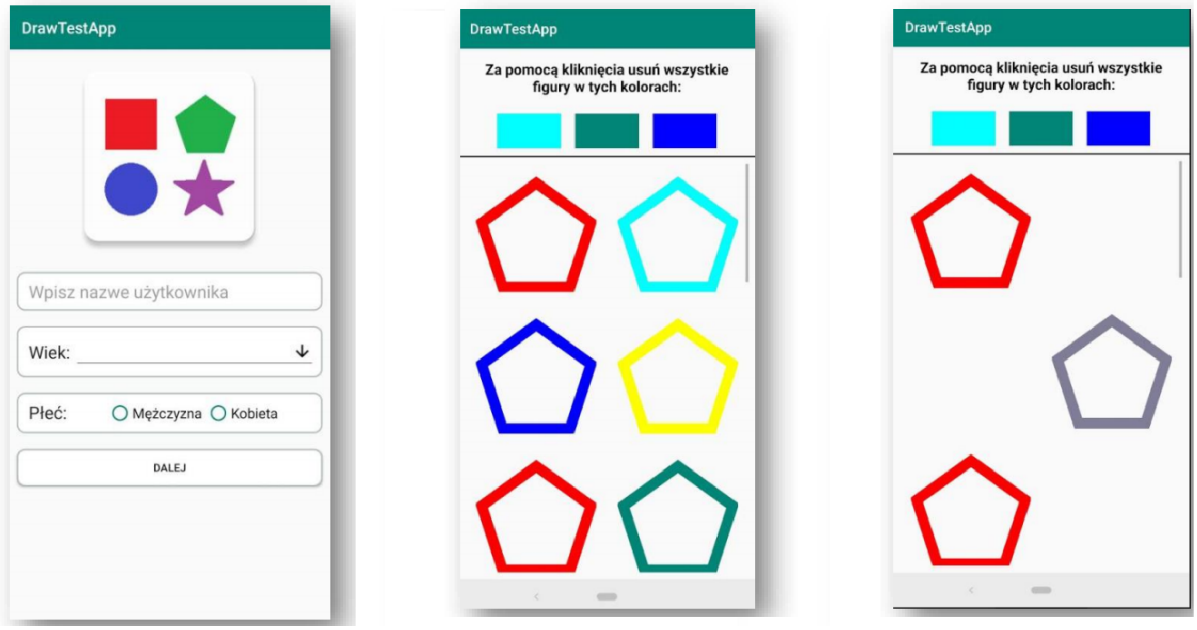


Figure D.1. Design of the application for the data collection. Source: screenshot from application developed by W. Wąsowska and P. Wojciechowski

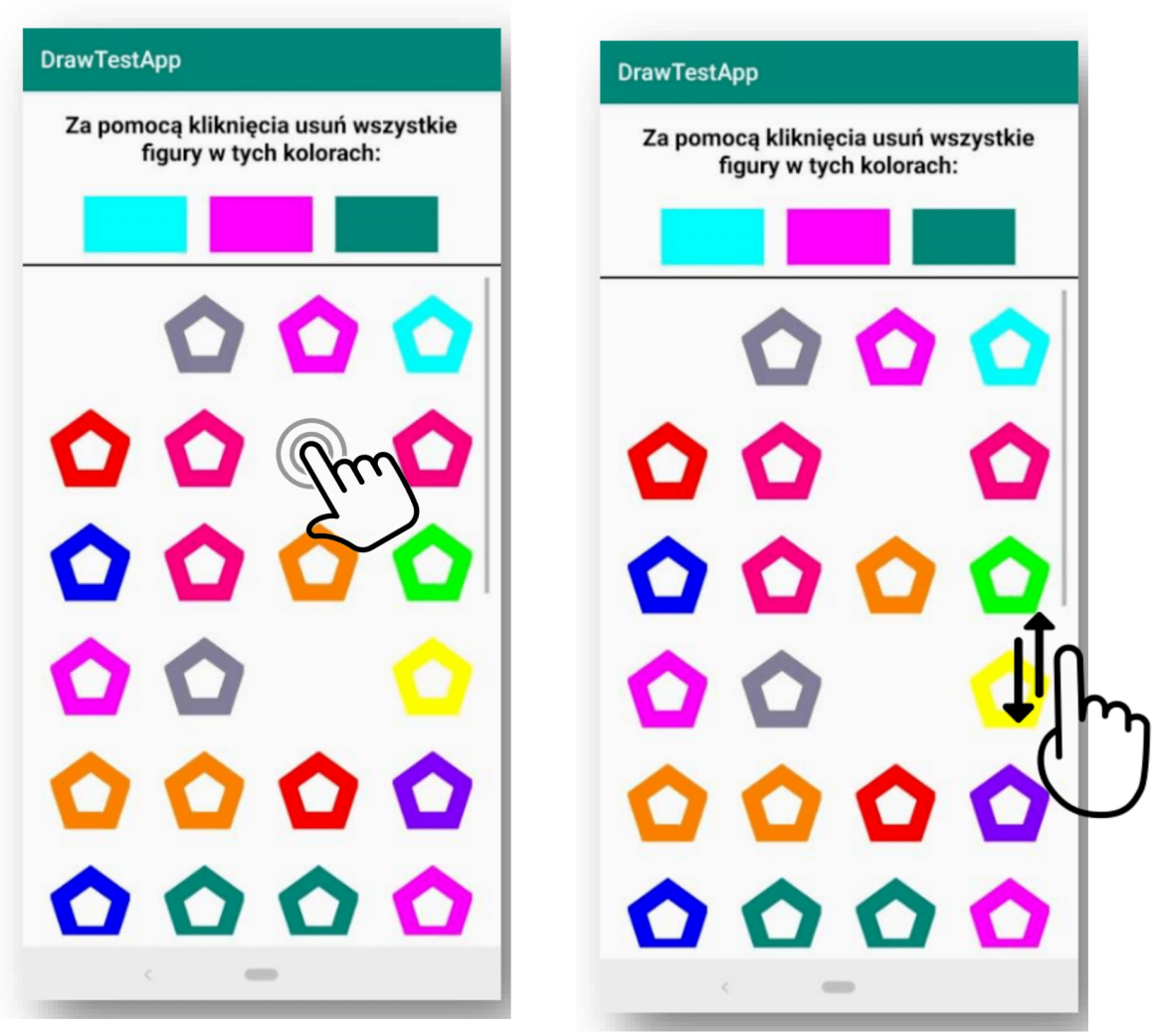


Figure D.2. Actions performed in the application for the data collection. Source: screenshot from application developed by W. Wąsowska and P. Wojciechowski

Appendix E

Authentication scenario presentation - detailed results

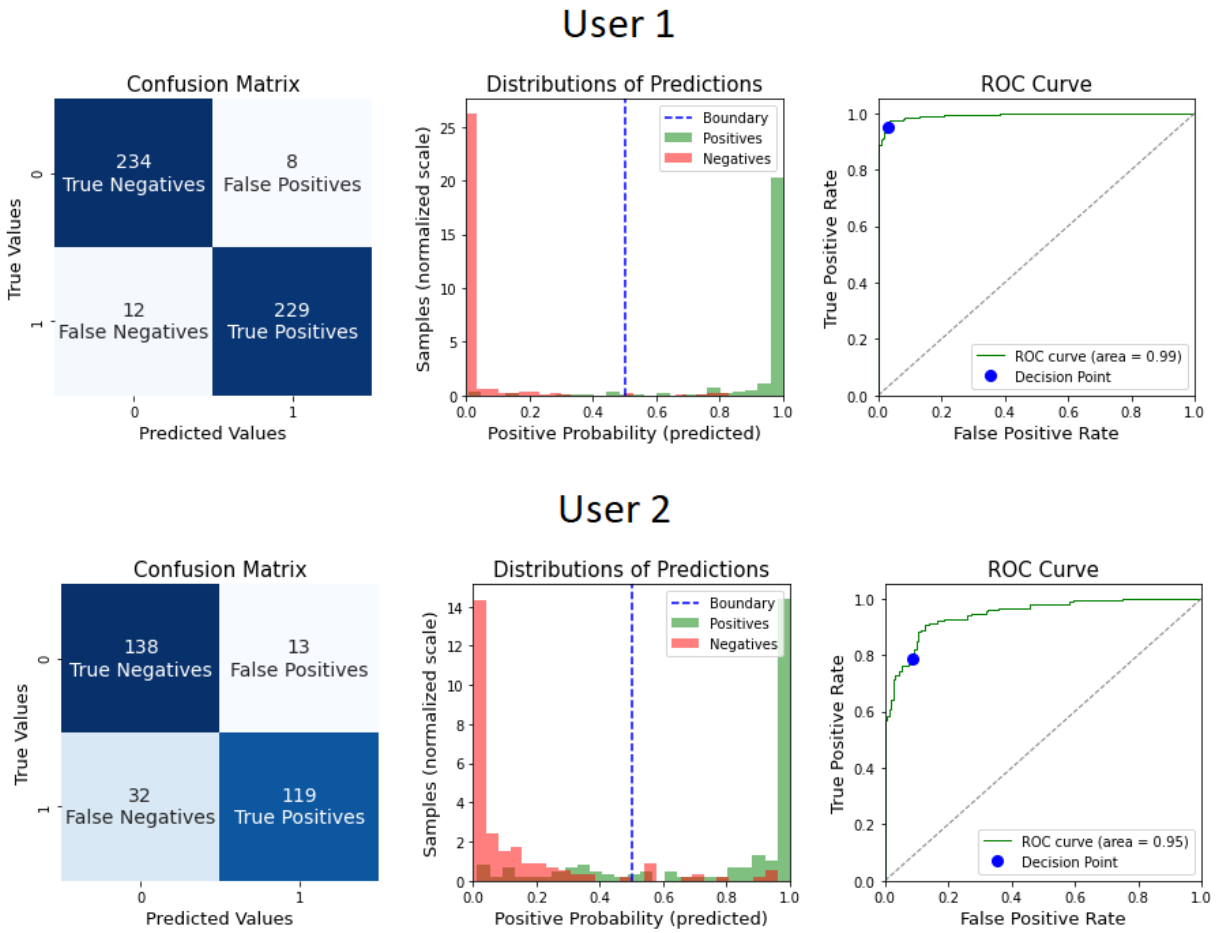


Figure E.1. Examples of evaluation criteria for two users in the authentication scenario - Touchalytics dataset. Source: own elaboration

Table E.1. Comparison of average method performance and macro averages in an authentication scenario with 2:1 impostor data ratio.

Dataset / Characteristic	Touchalytics	Sapienta BioIdent	Serwadda	Own Dataset
Accuracy	94,28%	95,25%	94,92%	92,49%
Precision	95,89%	97,51%	96,51%	93,90%
Recall	95,57%	95,32%	95,92%	95,22%
F-Score	95,71%	96,39%	96,18%	94,37%
EER	5,91%	4,71%	5,07%	7,50%
N=3 Accuracy	98,60%	99,22%	99,04%	98,14%
N=3 Precision	99,08%	99,42%	99,38%	99,30%
N=3 Recall	98,89%	99,47%	99,16%	98,15%
N=3 F-score	98,98%	99,44%	99,25%	98,41%
N=3 EER	0,55%	0,43%	0,34%	<0,01%
N=5 Accuracy	99,67%	99,83%	99,65%	94,06%
N=5 Precision	99,87%	99,91%	99,79%	99,65%
N=5 Recall	99,81%	99,91%	99,61%	99,72%
N=5 F-score	99,84%	99,91%	99,69%	99,85%
N=5 EER	0,02%	0,02%	0,04%	<0,01%

Source: own elaboration

Appendix F

Tap classification on the created dataset - detailed results

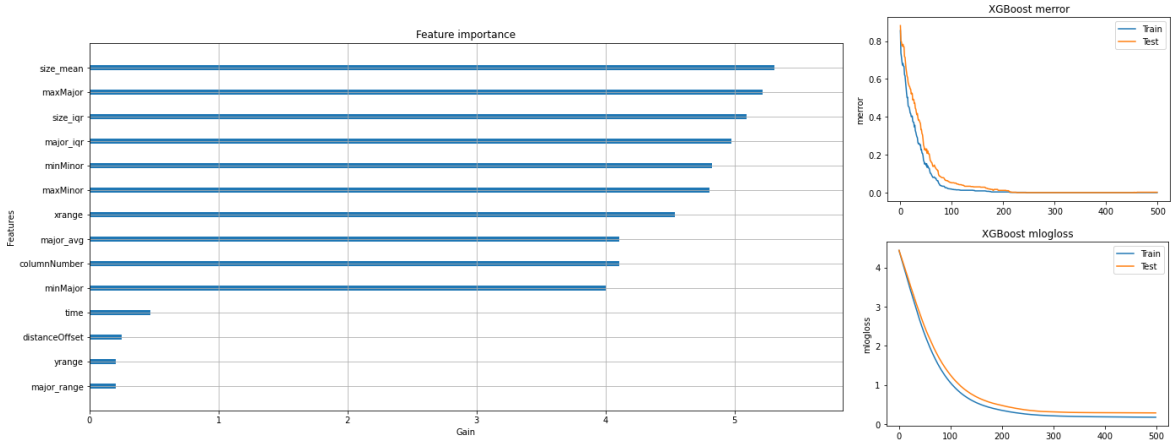


Figure F.1. Top feature importance (Gain) and the learning curves for the tap classification in the own dataset. Source: own elaboration

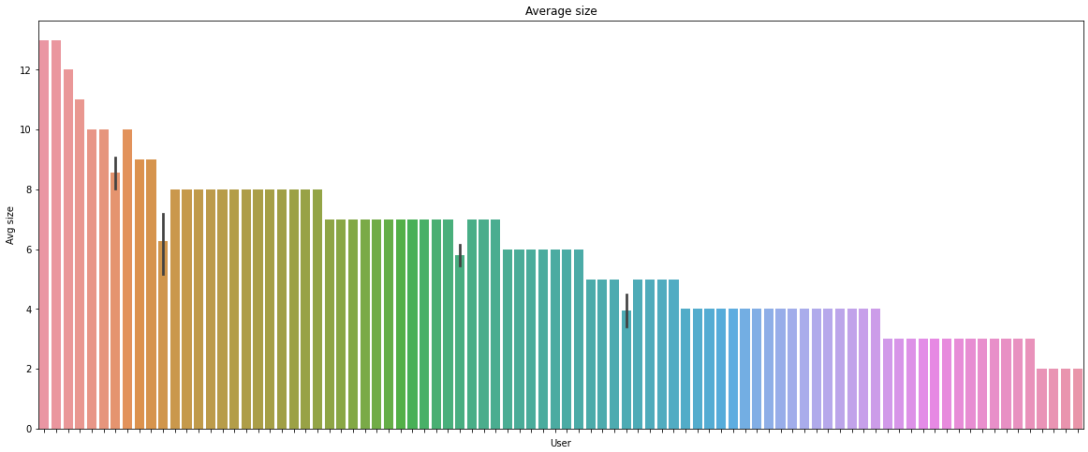


Figure F.2. Distribution of the mean touch size of a user for taps in the own Dataset. Source: own elaboration

Appendix G

Gender and age recognition - detailed results

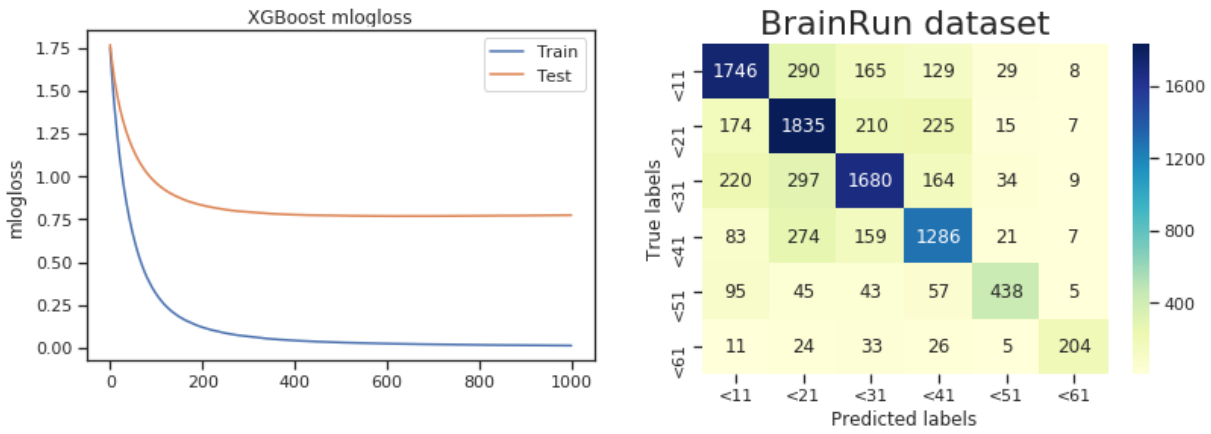


Figure G.1. Mlogloss during the process of classifier learning compared to the test dataset and nominal classification confusion matrix. Source: own elaboration

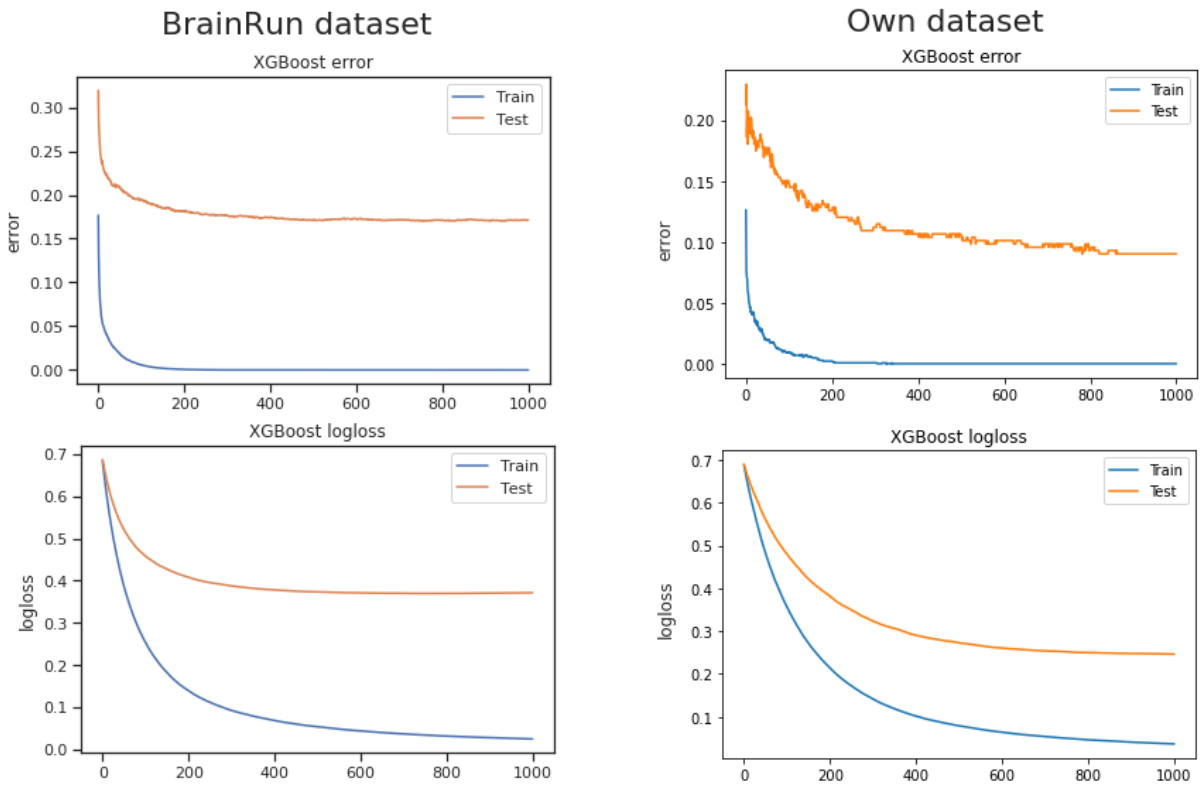


Figure G.2. Errors showcasing the learning process for gender recognition on both datasets. Own dataset on the left, BrainRun on the right. Source: own elaboration

Appendix H

Bank applications UI comparison

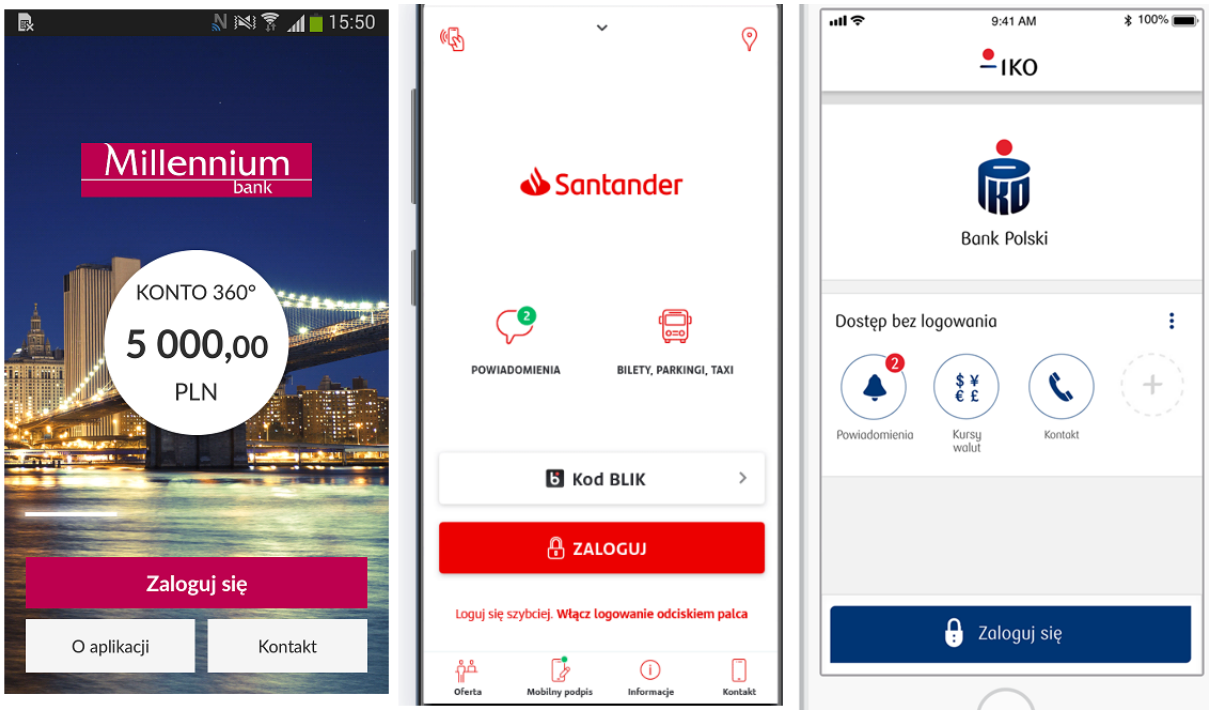


Figure H.1. Login page for the chosen 3 mobile bank applications. Source: screenshots of Polish mobile banking applications

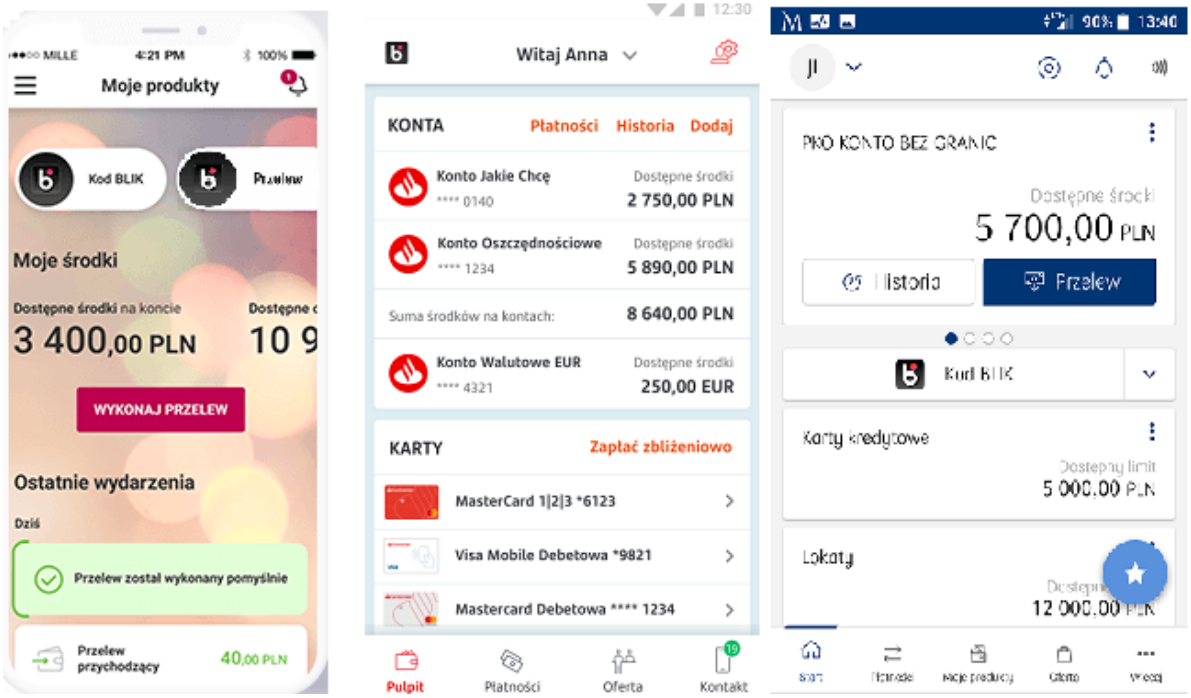


Figure H.2. Main menu for the chosen 3 mobile bank applications. Source: screenshots of Polish mobile banking applications

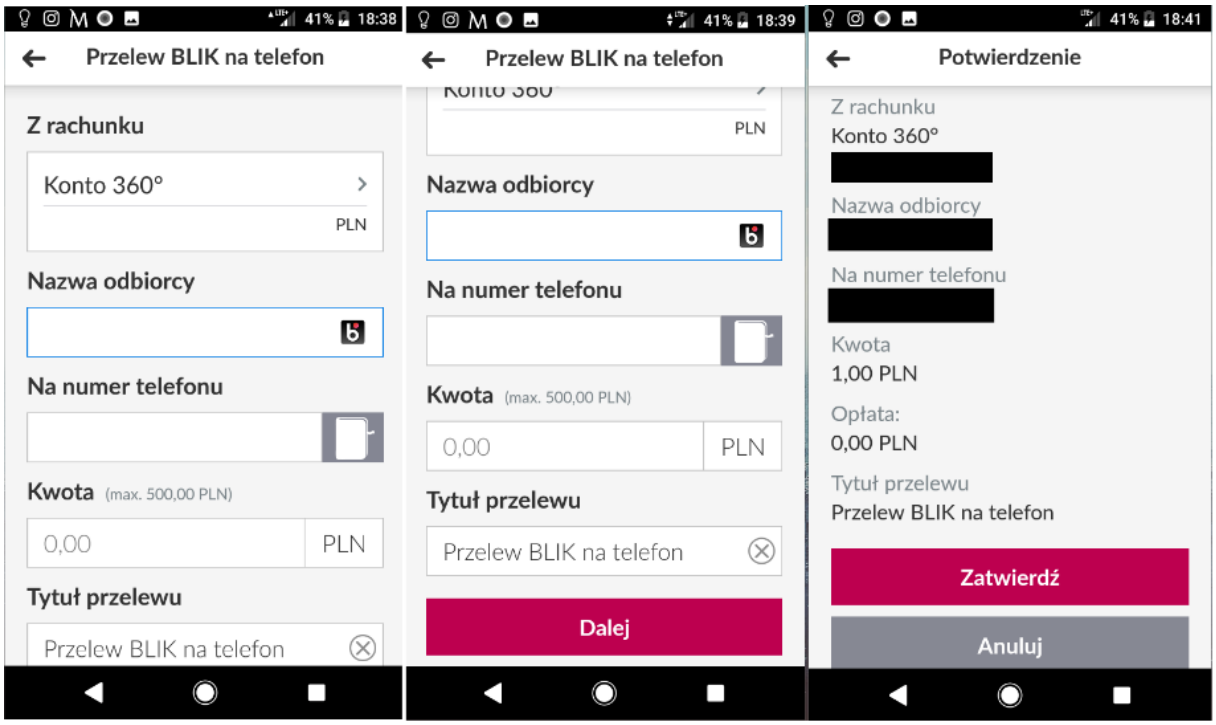


Figure H.3. Example of the money transfer process carried out by the Millenium Bank mobile application. Source: screenshots from Millenium bank application

Appendix I

Alternative UI designs

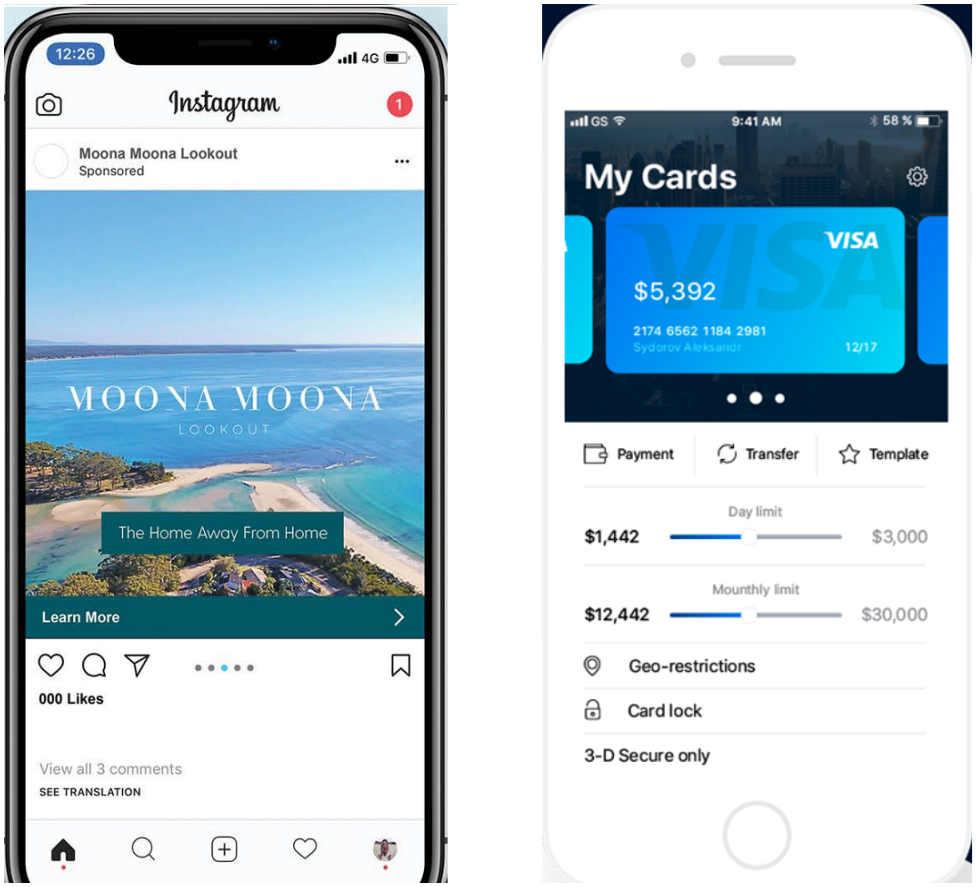


Figure I.1. UI carousel element on Instagram (left) and potential application in the financial application (right). Source: screenshots from Instagram application and (Syodorov, 2020)

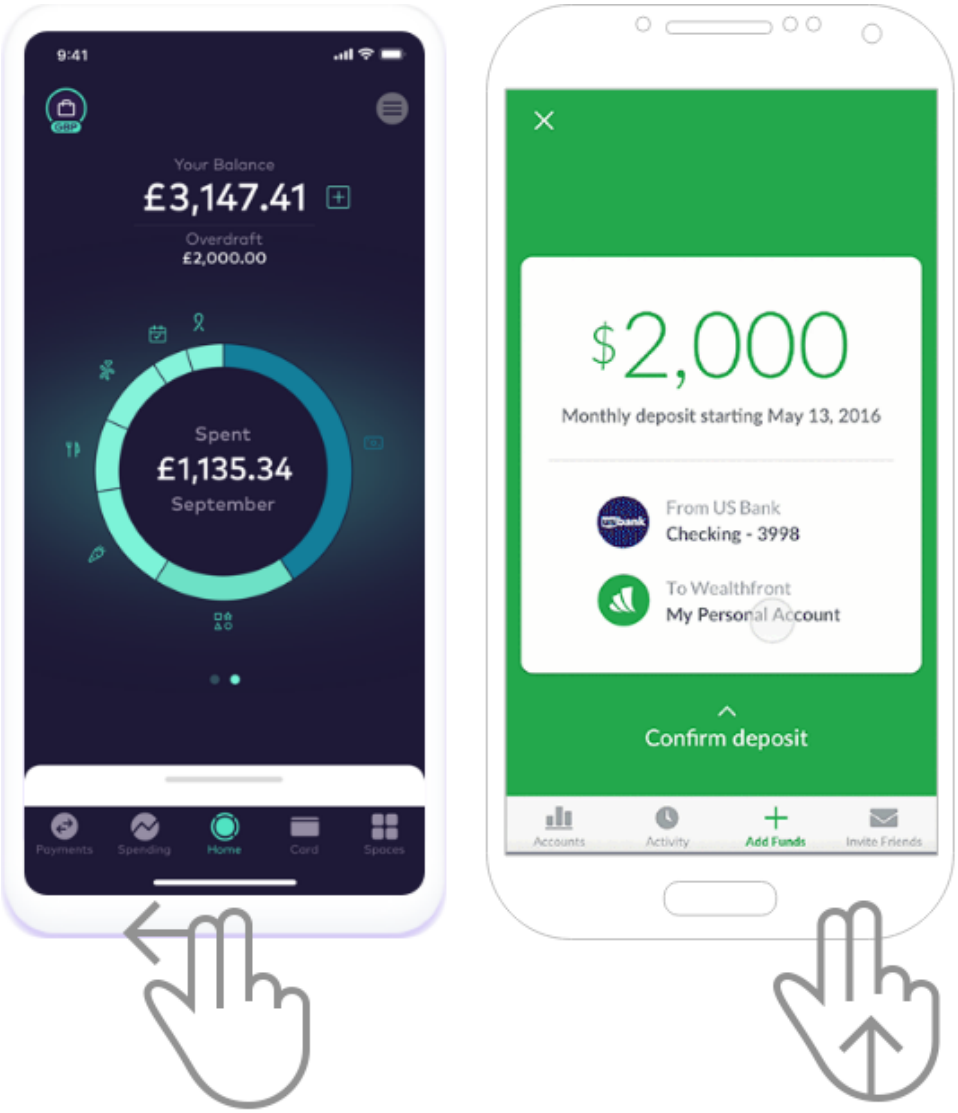


Figure I.2. Alternative UI design elements. Source: (DL Accounts Ltd, 2020; Sang, 2020)

Bibliography

Abdulhak, S. A., & Abdulaziz, A. A. (2018). A systematic review of features identification and extraction for behavioral biometrie authentication in touchscreen mobile devices. *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 68–73.

Abuhamad, M., Abusnaina, A., Nyang, D., & Mohaisen, D. (2020). Sensor-based Continuous Authentication of Smartphones' Users Using Behavioral Biometrics: A Contemporary Survey. *IEEE Internet of Things Journal*, *8*(1), 65–84.

Ali, M. L., Monaco, J. V., Tappert, C. C., & Qiu, M. (2017). Keystroke biometric systems for user authentication. *Journal of Signal Processing Systems*, *86*(2-3), 175–190.

Alonso-Fernandez, F., Bigun, J., Fierrez, J., Fronthaler, H., Kollreider, K., & Ortega-Garcia, J. (2009). Fingerprint recognition. *Guide to biometric reference systems and performance evaluation* (pp. 51–88). Springer.

Alotaibi, S., Furnell, S., & Clarke, N. (2015). Transparent authentication systems for mobile device security: A review. *Internet Technology and Secured Transactions (ICITST), 2015 10th International Conference for*, 406–413.

Alzubaidi, A., & Kalita, J. (2016). Authentication of smartphone users using behavioral biometrics. *IEEE Communications Surveys & Tutorials*, *18*(3), 1998–2026.

Antal, M., Bokor, Z., & Szabó, L. Z. (2015). Information revealed from scrolling interactions on mobile devices. *Pattern Recognition Letters*, *56*, 7–13.

Antal, M., Szabó, L. Z., & Bokor, Z. (2014). Identity Information Revealed From Mobile Touch Gestures. *Studia Universitatis Babes-Bolyai, Informatica*, *59*.

Awad, A. (2017). Collective Framework for Fraud Detection Using Behavioral Biometrics. *Information Security Practices* (pp. 29–37). Springer.

Axente, M.-S., Dobre, C., Ciobanu, R.-I., & Purnichescu-Purtan, R. (2020). Gait Recognition as an Authentication Method for Mobile Devices. *Sensors*, *20*(15), 4110.

Ayed, M. B. (2014). *Method for adaptive authentication using a mobile device* [US Patent 8,646,060].

Bailey, K. O., Okolica, J. S., & Peterson, G. L. (2014). User identification and authentication using multi-modal behavioral biometrics. *Computers & Security*, *43*, 77–89.

Baqeel, H., & Saeed, S. (2019). Face Detection Authentication on Smartphones: End Users Usability Assessment Experiences. *2019 International Conference on Computer and Information Sciences (ICCIS)*, 1–6.

Błach, J. (2011). Financial innovations and their role in the modern financial system-identification and systematization of the problem. *e-Finanse: Financial Internet Quarterly*, *7*(3), 13–26.

Black, P., Gondal, I., & Layton, R. (2018). A survey of similarities in banking malware behaviours. *Computers & Security*, *77*, 756–772.

Bo, C., Zhang, L., Li, X.-Y., Huang, Q., & Wang, Y. (2013). Silentsense: silent user identification via touch and movement behavioral biometrics. *Proceedings of the 19th annual international conference on Mobile computing & networking*, 187–190.

Boczoń, W. (2017). *Biometria w bankowości. Co za jej pomocą załatwimy dziś w banku?* [Online; Accessed 10.09.2020]. https://www.bankier.pl/wiadomosc/Biometria-w-bankowosci-Co-za-jej-pomoca-zalatwimy-dzis-w-banku-7542743.html

Boczoń, W. (2020). *Raport PRNews.pl: Liczba klientów mobile only – IV kw. 2020*.

Bohne, J. (2018). *Beyond the ROC AUC: Toward Defining Better Performance Metrics* [Online; Accessed 10.02.2020]. https://medium.com/bcggamma/beyond-the-roc-auc-toward-defining-better-performance-metrics-b11f5d35adda

Bolesławski, M., & Nowakowski, E. W. (2016). Innowacje bankowe a bezpieczeństwo systemu finansowego w Polsce. *Journal of Modern Science*, (3 (30)), 263–281.

Bolle, R. M., Connell, J. H., Pankanti, S., Ratha, N. K., & Senior, A. W. (2013). *Guide to biometrics*. Springer Science & Business Media.

Bonneau, J., & Preibusch, S. (2010). The Password Thicket: Technical and Market Failures in Human Authentication on the Web. *WEIS*.

Bonneau, J., Preibusch, S., & Anderson, R. (2012). A birthday present every eleven wallets? The security of customer-chosen banking PINs. *International Conference on Financial Cryptography and Data Security*, 25–40.

Boukerche, A., & Notare, M. S. M. A. (2002). Behavior-based intrusion detection in mobile phone systems. *Journal of Parallel and Distributed Computing*, *62*(9), 1476–1490.

Buriro, A., Crispo, B., Del Frari, F., Klardie, J., & Wrona, K. (2016). ITSME: Multi-modal and Unobtrusive Behavioural User Authentication for Smartphones. *Technology and Practice of Passwords*, 45.

Bursztein, E. (2014). *Survey: Most people don't lock their Android phones — but should* [Online; Accessed: 08.09.2020]. https://www.elie.net/blog/survey-most-people-dont-lock-their-android-phones-but-should

Buschkes, R., Kesdogan, D., & Reichl, P. (1998). How to increase security in mobile networks by anomaly detection. *Computer Security Applications Conference, 1998. Proceedings. 14th Annual*, 3–12.

Cao, K., & Jain, A. K. (2016). Hacking Mobile Phones Using 2D Printed Fingerprints.

Chatterjee, K. et al. (n.d.). Continuous User Authentication System: A Risk Analysis Based Approach. *Wireless Personal Communications*, 1–15.

Chen, S., Lach, J., Lo, B., & Yang, G.-Z. (2016). Toward pervasive gait analysis with wearable sensors: A systematic review. *IEEE journal of biomedical and health informatics*, *20*(6), 1521–1537.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Choi, S., Chang, I., & Teoh, A. B. J. (2018). One-class Random Maxout Probabilistic Network for Mobile Touchstroke Authentication. *2018 24th International Conference on Pattern Recognition (ICPR)*, 3359–3364.

Crawford, H., & Renaud, K. (2014). Understanding user perceptions of transparent authentication on a mobile device. *Journal of Trust Management*, *1*(1), 7.

Czyżewski, A., Hoffmann, P., Szczuko, P., Kurowski, A., Lech, M., & Szczodrak, M. (2019). Analysis of results of large-scale multimodal biometric identity verification experiment. *IET Biometrics*, *8*(1), 92–100.

Damaševičius, R., Maskeliūnas, R., Venčkauskas, A., & Woźniak, M. (2016). Smartphone user identity verification using gait characteristics. *Symmetry*, *8*(10), 100.

Damopoulos, D., Kambourakis, G., & Gritzalis, S. (2013). From keyloggers to touchloggers: Take the rough with the smooth. *Computers & security*, *32*, 102–114.

Datagenetics.com. (2012). *PIN analysis* [Online; Accessed: 2020-10-10]. http://www.datagenetics.com/blog/september32012/

Deloitte Center for Financial Services. (2018). *2018 Banking Outlook* [Online; Accessed 10.10.2019]. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/gx-fsi-dcfs-2018-banking-outlook.pdf

Derawi, M. O., Bours, P., & Holien, K. (2010). Improved cycle detection for accelerometer based gait authentication. *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 312–317.

Diep, N. N., Pham, C., & Phuong, T. M. (2015). SigVer3D: Accelerometer Based Verification of 3-D Signatures on Mobile Devices. *Knowledge and Systems Engineering* (pp. 353–365). Springer.

DL Accounts Ltd. (2020). *Starling bank UI example* [Online; Accessed 10.09.2020]. https://www.dlaccounts.co.uk/partners/starling-bank/

Doel, K. (2015). *SplashData's fifth annual "Worst Passwords List"* [Accessed: 2020-08-09]. https://www.teamsid.com/worst-passwords-2015/

EBA, E. B. A. (2019). *Opinion of the European Banking Authority on the elements of strong customer authentication under PSD2* [Online; Accessed 10.10.2019]. https://ec.europa.eu/info/publications/190621-eba-opinion-strong-customer-authentication_en

European Central Bank. (2019). *European Central Bank, 2018. Fifth report on card fraud*. https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport202008~521edb602b.en.html

Europen Comission. (2015). *PSD 2, Directive (EU) 2015/2366*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32015L2366

Feng, T., Liu, Z., Kwon, K.-A., Shi, W., Carbunar, B., Jiang, Y., & Nguyen, N. (2012). Continuous mobile authentication using touchscreen gestures. *2012 IEEE Conference on Technologies for Homeland Security (HST)*, 451–456.

Feng, T., Yang, J., Yan, Z., Tapia, E. M., & Shi, W. (2014). Tips: Context-aware implicit user identification using touch screen in uncontrolled environments. *Proceedings of the 15th Workshop on Mobile Computing Systems and Applications*, 9.

Ficowicz, G. (2018). Biometrics – Discard Your Fear of the Unknown. *Wyzwania Informatyki Bankowej 2018*, (5), 267–276.

Fierrez, J., Pozo, A., Martinez-Diaz, M., Galbally, J., & Morales, A. (2018). Benchmarking touchscreen biometrics for mobile authentication. *IEEE Transactions on Information Forensics and Security*, *13*(11), 2720–2733.

Financial Stability Board. (2020). *Global Monitoring Report on Non-Bank Financial Intermediation 2020*. https://www.fsb.org/2020/12/global-monitoring-report-on-non-bank-financial-intermediation-2020/

Fingerprint Unlock Security: iOS vs. Google Android (Part II) [Online; Accessed 10.09.2020]. (2016). http://blog.elcomsoft.com/2016/06/fingerprint-unlock-security-ios-vs-google-android-part-ii/

Frank, M., Biedert, R., Ma, E., Martinovic, I., & Song, D. (2012). Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE transactions on information forensics and security*, *8*(1), 136–148.

Fridman, L., Weber, S., Greenstadt, R., & Kam, M. (2015). Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location.

Fridman, L., Weber, S., Greenstadt, R., & Kam, M. (2017). Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location. *IEEE Systems Journal*, *11*(2), 513–521.

Gadzina, B. (2020). *Santander App redesign concept- BLIK code payment* [Online; Accessed 10.08.2020]. https://dribbble.com/shots/8328900

Gascon, H., Uellenbeck, S., Wolf, C., & Rieck, K. (2014). Continuous Authentication on Mobile Devices by Analysis of Typing Motion Behavior. *Sicherheit*, 1–12.

Giuffrida, C., Majdanik, K., Conti, M., & Bos, H. (2014). I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics. *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, 92–111.

Goguey, A., Casiez, G., Cockburn, A., & Gutwin, C. (2018). Storyboard-based empirical modeling of touch interface performance. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, *37*(2).

GSMA. (2019). *The Mobile Economy Report*. https://www.gsma.com/r/mobileeconomy/

gs.statcounter.com. (2020). *iOS market share worldwide* [Online; Accessed 12.07.2020]. https://gs.statcounter.com/os-market-share/mobile/worldwide

Guerra-Casanova, J., Sánchez-Ávila, C., Bailador, G., & de Santos Sierra, A. (2012). Authentication in mobile devices through hand gesture recognition. *International Journal of Information Security*, *11*(2), 65–83.

Günther, M., El Shafey, L., & Marcel, S. (2016). Face Recognition in Challenging Environments: An Experimental and Reproducible Research Survey. *Face Recognition Across the Imaging Spectrum* (pp. 247–280). Springer.

Hayashi, E., Riva, O., Strauss, K., Brush, A., & Schechter, S. (2012). Goldilocks and the two mobile devices: going beyond all-or-nothing access to a device's applications. *Proceedings of the Eighth Symposium on Usable Privacy and Security*, 2.

Hernández-Álvarez, L., de Fuentes, J. M., González-Manzano, L., & Hernández Encinas, L. (2021). Privacy-Preserving Sensor-Based Continuous Authentication and User Profiling: A Review. *Sensors*, *21*(1), 92.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems*, *19*(2), 4.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–105.

Hilas, C. S., & Sahalos, J. N. (2005). User profiling for fraud detection in telecommunication networks. *5th International Conference on Technology and Automation*, 382–387.

Imgraben, J., Engelbrecht, A., & Choo, K.-K. R. (2014). Always connected, but are smart mobile users getting more security savvy? A survey of smart mobile device users. *Behaviour & Information Technology*, *33*(12), 1347–1360.

Inoue, M., & Ogawa, T. (2018). TapOnce: a novel authentication method on smartphones. *International Journal of Pervasive Computing and Communications*.

Jain, A., Hong, L., & Pankanti, S. (2000). Biometric identification. *Communications of the ACM*, *43*(2), 90–98.

Jain, A., & Kanhangad, V. (2015). Exploring orientation and accelerometer sensor data for personal authentication in smartphones using touchscreen gestures. *Pattern recognition letters*, *68*, 351–360.

Jain, A., & Kanhangad, V. (2019). Gender recognition in smartphones using touchscreen gestures. *Pattern Recognition Letters*, *125*, 604–611.

Kainda, R., Flechais, I., & Roscoe, A. (2010). Security and usability: Analysis and evaluation. *2010 International Conference on Availability, Reliability and Security*, 275–282.

Kałużny, P. (2017). Behavioural profiling authentication based on trajectory based anomaly detection model of user's mobility. *International Conference on Business Information Systems*, 242–254.

Kałużny, P. (2019a). Behavioral Biometrics in Mobile Banking and Payment Applications. In W. Abramowicz & A. Paschke (Eds.), *BIS 2018 International Workshops, Berlin, Germany, July 18–20, 2018, Revised Papers* (pp. 646–658). https://doi.org/10.1007/978-3-030-04849-5_55

Kałużny, P. (2019b). Touchscreen behavioural biometrics authentication in self-contained mobile applications design. In W. Abramowicz & A. Paschke (Eds.), *International Conference on Business Information Systems* (pp. 672–685).

Kałużny, P., & Filipowska, A. (2018). Large Scale Mobility-based Behavioral Biometrics on the Example of the Trajectory-based Model for Anomaly Detection. *j jucs*, *24*, 417–443.

Kałużny, P., & Stolarski, P. (2019). Biometria behawioralna i „tradycyjna" w mobilnych usługach bankowych – stan oraz przyszłe możliwości zastosowania. *Bezpieczny Bank*, (1), 139–161. https://doi.org/10.26354/bb.7.1.74.2019

Kayacik, H. G., Just, M., Baillie, L., Aspinall, D., & Micallef, N. (2014). Data driven authentication: On the effectiveness of user behaviour modelling with mobile device sensors. *arXiv preprint arXiv 1410 7743*.

Kecman, V. (2005). Support vector machines–an introduction. *Support vector machines: theory and applications* (pp. 1–47). Springer.

Kindt, E. J. (2013). An Introduction into the Use of Biometric Technology. *Privacy and Data Protection Issues of Biometric Applications* (pp. 15–85). Springer.

KNF. (2020). *Aktywa Sektora Bankowego.* [Online; Accessed 02.02.2020]. https://www.nbp.pl/home.aspx?f=/statystyka/pieniezna_i_bankowa/naleznosci.html

Krenker, A., Volk, M., Sedlar, U., Bešter, J., & Kos, A. (2009). Bidirectional artificial neural networks for mobile-phone fraud detection. *Etri Journal*, *31*(1), 92–94.

Kumar, R., Kundu, P. P., & Phoha, V. V. (2018). Continuous authentication using one-class classifiers and their fusion. *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, 1–8.

Lamiche, I., Bin, G., Jing, Y., Yu, Z., & Hadid, A. (2019). A continuous smartphone authentication method based on gait patterns and keystroke dynamics. *Journal of Ambient Intelligence and Humanized Computing*, *10*(11), 4417–4430.

Lawless Research. (2016). *Beyond the Password: The Future of Account Security* [Online; Accessed 01.10.2020]. https://www.telesign.com/wp-content/uploads/2016/06/Telesign-Report-Beyond-the-Password-June-2016-1.pdf

Li, F., Clarke, N., Papadaki, M., & Dowland, P. (2011). Behaviour profiling for transparent authentication for mobile devices. *European Conference on Cyber Warfare and Security*, 307.

Li, F., Clarke, N., Papadaki, M., & Dowland, P. (2014). Active authentication for mobile devices utilising behaviour profiling. *International journal of information security*, *13*(3), 229–244.

Li, L., Zhao, X., & Xue, G. (2013). Unobservable re-authentication for smartphones. *NDSS*, 1–16.

Lord, N. (2018). *Uncovering Password Habits: Are Users' Password Security Habits Improving?* [Online; Accessed 10.02.2020]. https://digitalguardian.com/blog/uncovering-password-habits-are-users-password-security-habits-improving-infographic

Lovisotto, G., Malik, R., Sluganovic, I., Roeschlin, M., Trueman, P., & Martinovic, I. (2017). Mobile biometrics in financial services: A five factor framework. *University of Oxford, Oxford, UK*.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates, Inc.

Luo, X., Li, H., Zhang, J., & Shim, J. P. (2010). Examining multi-dimensional trust and multi-faceted risk in initial acceptance of emerging technologies: An empirical study of mobile banking services. *Decision support systems*, *49*(2), 222–234.

Mahbub, U., Sarkar, S., Patel, V. M., & Chellappa, R. (2016). Active user authentication for smartphones: A challenge data set and benchmark results. *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 1–8.

Marcinkowska, M. (2012). Innowacje finansowe w bankach. *Acta Universitatis Lodziensis. Folia Oeconomica*, *266*, 71–96.

Mason, J. E., Traore, I., & Woungang, I. (2019). Facets and Promises of Gait Biometric Recognition. *Biometric-Based Physical and Cybersecurity Systems* (pp. 233–253). Springer.

Masood, R., Zhao, B. Z. H., Asghar, H. J., & Kaafar, M. A. (2018). Touch and you're trapp (ck) ed: Quantifying the uniqueness of touch gestures for tracking. *Proceedings on Privacy Enhancing Technologies*, *2018*(2), 122–142.

Mastercard. (2017). *Raport Mastercard Bankowość mobilna - trendy i wyróżniki oferty w Polsce i na świecie* [Online; Accessed 10.09.2018]. http://konferencje.alebank.pl/wp-content/uploads/2017/06/PM.Bankowosc-mobilna.Adam_.Splawski.Mastercard.pdf

Mastercard. (2018). *Mastercard Market Intelligence Report "Biometrics - Meeting the challenge of authentication and payments technology"* [Online; Accessed 10.10.2019]. https://www.mastercard.us/content/dam/public/mastercardcom/na/us/en/smb/other/biometrics_updated_030619.pdf

Mazhelis, O. (2007). One-class classifiers: a review and analysis of suitability in the context of mobile-masquerader detection. *Arima Journal*, *6*, 29–48.

Mazhelis, O., & Puuronen, S. (2007). A framework for behavior-based detection of user substitution in a mobile context. *computers & security*, *26*(2), 154–176.

MBank. (2020). *Wirtualna Ekarta płatnicza* [Online; Accessed 10.08.2020]. https://www.mbank.pl/private-banking/uslugi-bankowe/karty-do-konta/ekarta/

McDonnell, C., Fox, J., Moroney, M., & Wills, S. (2014). *Mobile devices Secure or security risk?* [Online; Accessed 10.09.2020]. http://www2.deloitte.com/content/dam/Deloitte/ie/Documents/Risk/mobile_device_secure_security_risk.pdf

Mecke, L., Prange, S., Buschek, D., & Alt, F. (2018). A Design Space for Security Indicators for Behavioural Biometrics on Mobile Touchscreen Devices. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, LBW003.

Meng, Y., Wong, D. S., Schlegel, R., et al. (2012). Touch gestures based biometric authentication scheme for touchscreen mobile phones. *International Conference on Information Security and Cryptology*, 331–350.

Meola, A. (2019). *The digital trends disrupting the banking industry in 2019*. https://www.businessinsider.com/banking-industry-trends?IR=T

Mills, A. M., & Zheng, Z. (2019). The Future of Identity Management: Understanding Consumer Attitudes Towards Biometric Identification.

Milton, L. C., & Memon, A. (2016). Intruder detector: A continuous authentication tool to model user behavior. *Intelligence and Security Informatics (ISI), 2016 IEEE Conference on*, 286–291.

Morgan, R. (2017). *Fintech trends driving the new decade* [Online; Accessed 07.12.2018]. https://bankingjournal.aba.com/2017/09/the-top-fintech-trends-driving-the-next-decade/

Mulders, M., & den Hertog, P. (2003). Measuring innovative behaviour in Dutch Financial Services: a meso perspective. *SIID-project, phase*, *4*.

Muslukhov, I., Boshmaf, Y., Kuo, C., Lester, J., & Beznosov, K. (2013). Know your enemy: the risk of unauthorized access in smartphones by insiders. *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, 271–280.

Narodowy Bank Polski. (2020). *Informacja o rozliczeniach i rozrachunkach międzybankowych w III kwartale 2019 r.* [Online; Accessed 12.03.2020]. https://www.nbp.pl/systemplatniczy/publikacje/2019_3.pdf

National Institute of Standards and Technology. (2014). *Hacker fakes German minister's fingerprints using photos of her hands* [Online; Accessed 10.02.2020]. https://www.theguardian.com/technology/2014/dec/30/hacker-fakes-german-ministers-fingerprints-using-photos-of-her-hands

National Institute of Standards and Technology. (2020). *Face Recognition Vendor Test (FRVT)* [Online; Accessed 10.02.2020]. https://pages.nist.gov/frvt/reports/1N/frvt_1N_report.pd

Ngo, T. T., Makihara, Y., Nagahara, H., Mukaigawa, Y., & Yagi, Y. (2014). The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition*, *47*(1), 228–237.

Olszak, C. (2007). Wyzwania ery wiedzy,[w:] CM Olszak, E. *Strategie i modele gospodarki elektronicznej Wydawnictwo Naukowe PWN Warszawa*.

Panchumarthy, R., Subramanian, R., & Sarkar, S. (2012). Biometric evaluation on the cloud: A case study with humanid gait challenge. *2012 IEEE Fifth International Conference on Utility and Cloud Computing*, 219–222.

Papamichail, M. D., Chatzidimitriou, K. C., Karanikiotis, T., Oikonomou, N.-C. I., Symeonidis, A. L., & Saripalle, S. K. (2019). BrainRun: A Behavioral Biometrics Dataset towards Continuous Implicit Authentication. *Data*, *4*(2), 60.

Patel, V. M., Chellappa, R., Chandra, D., & Barbello, B. (2016). Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine*, *33*(4), 49–61.

Pavlovic, M., Petrovic, R., Stojanovic, B., & Stankovic, S. (2018). Facial expression and lighting conditions influence on face recognition performance. *Proceedings of the 5th International Conference IcETRAN*, 777–781.

Polowczyk, J. (2009). Podstawy ekonomii behawioralnej. *Przegląd organizacji*, (12), 3–7.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2014). Artifact Evaluation in Information Systems Design-Science Research-a Holistic View. *PACIS*, 23.

Primo, A., Phoha, V. V., Kumar, R., & Serwadda, A. (2014). Context-aware active authentication using smartphone accelerometer measurements. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 98–105.

PwC. (2017). *Visa - Digital Payments Study 2017 Germany* [Online; Accessed 10.09.2020]. https://www.pwc.de/mobilepayment

Ramachandra, R., Venkatesh, S., Raja, K. B., Bhattacharjee, S., Wasnik, P., Marcel, S., & Busch, C. (2019). Custom silicone face masks: Vulnerability of commercial face recognition systems & presentation attack detection. *2019 7th International Workshop on Biometrics and Forensics (IWBF)*, 1–6.

Rattani, A., & Derakhshani, R. (2018). A survey of mobile face biometrics. *Computers & Electrical Engineering*, *72*, 39–52.

Reid, P. (2004). *Biometrics for network security*. Prentice Hall Professional.

Renaud, K. (2005). Evaluating authentication mechanisms. *Security and usability*, 103–128.

Revolut. (2021). *Revolut - virtual cards* [Online; Accessed 27.03.2021]. https://blog.revolut.com/content/images/2018/03/Disposable-Virtual-Card.gif

Saeed, K. (2012). Biometrics principles and important concerns. *Biometrics and Kansei Engineering* (pp. 3–20). Springer.

Saevanee, H., & Bhattarakosol, P. (2009). Authenticating user using keystroke dynamics and finger pressure. *2009 6th IEEE Consumer Communications and Networking Conference*, 1–2.

Saevanee, H., Clarke, N., Furnell, S., & Biscione, V. (2014). Text-based active authentication for mobile devices. *IFIP International Information Security Conference*, 99–112.

Saevanee, H., Clarke, N. L., & Furnell, S. M. (2012a). Multi-modal behavioural biometric authentication for mobile devices. *IFIP International Information Security Conference*, 465–474.

Saevanee, H., Clarke, N. L., & Furnell, S. M. (2012b). Multi-modal behavioural biometric authentication for mobile devices. *IFIP International Information Security Conference*, 465–474.

Samet, S., Ishraque, M. T., Ghadamyari, M., Kakadiya, K., Mistry, Y., & Nakkabi, Y. (2019). TouchMetric: a machine learning based continuous authentication feature testing mobile application. *International Journal of Information Technology*, 1–7.

Samojło, G. (2019). *How Mobile Apps Are Changing the Banking Industry: 5 Examples*. https://www.netguru.com/blog/mobile-apps-in-banking-examples

Samsung facial recgonition system "Face unlock" [Online; Accessed: 19.07.2020]. (2011). http://www.bloomberg.com/news/articles/2011-10-19/google-samsung-to-offer-phone-with-ice-cream-sandwich-software

Sang, S. (2020). *Swipe up and deposit UI* [Online; Accessed 10.09.2020]. https://dribbble.com/shots/3353747

Sanjith, R. (2017). Accelerating mobile biometrics adoption: doing nothing is not an option. *Biometric Technology Today*, *2017*(9), 5–9.

Serwadda, A., Phoha, V. V., & Wang, Z. (2013). Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms. *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, 1–8.

Shahzad, M., Liu, A. X., & Samuel, A. (2013). Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it. *Proceedings of the 19th annual international conference on Mobile computing & networking*, 39–50.

Shi, E., Niu, Y., Jakobsson, M., & Chow, R. (2010). Implicit authentication through learning user behavior. *International Conference on Information Security*, 99–113.

Singh, A., Chandrashekar, A., & Singh, V. (2020). *Banking Trends 2020* [Online; Accessed 10.07.2020]. https://www.capgemini.com/wp-content/uploads/2019/11/Retail_Banking_Trends_2020-1.pdf

Singha, T. B., Nath, R. K., & Narsimhadhan, A. (2017). Person recognition using smartphones' accelerometer data. *arXiv preprint arXiv:1711.04689*.

Sitová, Z., Šeděnka, J., Yang, Q., Peng, G., Zhou, G., Gasti, P., & Balagani, K. S. (2016). HMOG: New behavioral biometric features for continuous authentication of smartphone users. *IEEE Transactions on Information Forensics and Security*, *11*(5), 877–892.

Srinivas, V., Fromhart, S., Goradia, U., & Richa, W. (2018). *Banking Outlook "Accelerating the transformation"* [Online; Accessed 07.12.2018]. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/gx-fsi-dcfs-2018-banking-outlook.pdf

Staszczyk, M. (2016). Ochrona konsumentów korzystających z usług bankowości elektronicznej-na przykładzie ankiety przeprowadzonej wśród osób pracujących i/lub studiujących w Łodzi. *Bezpieczny Bank*, (1 (62)), 149–164.

Strohm, M. (2021). *Digital Banking Survey* [Online; Accessed 01.04.2021]. https://www.forbes.com/advisor/banking/digital-banking-survey-mobile-app-valuable-features/

Subasinghe, A. (2019). *Implementing a successful Open Banking Architecture* [Online; Accessed 10.02.2020]. https://wso2.com/articles/2019/08/implementing-a-successful-open-banking-architecture/

Sudoł, M., & Woszczyński, T. (2018). „Biometria w bankowości" – analiza i podsumowanie bieżącego statusu oraz roadmapa na najbliższą przyszłość. *Wyzwania Informatyki Bankowej 2018*, (5), 257–65.

Sultana, M., Paul, P. P., & Gavrilova, M. (2014). A concept of social behavioral biometrics: motivation, current developments, and future trends. *Cyberworlds (CW), 2014 International Conference on*, 271–278.

Świeszczak, K. (2017). Zaufanie w świecie finansów w obliczu rozwoju technologii na przykładzie banków i sektora FinTech. *Bezpieczny Bank*, (2 (67)), 143–158.

Syed, Z., Helmick, J., Banerjee, S., & Cukic, B. (2019). Touch gesture-based authentication on mobile devices: the effects of user posture, device size, configuration, and inter-session variability. *Journal of Systems and Software*, *149*, 158–173.

Syodorov, A. (2020). *Banking App free UI kit* [Online; Accessed 10.09.2020]. https://pinspiry.com/banking-app-free-ui-kit/

Szczekocka, E., Gromada, J., Filipowska, A., Jankowiak, P., Kałuzny, P., Brun, A., Portugal, J. M., & Staiano, J. (2016). Managing personal information: a telco perspective, 1–8.

Teh, P. S., Zhang, N., Teoh, A. B. J., & Chen, K. (2016). A survey on touch dynamics authentication in mobile devices. *Computers & Security*, *59*, 210–235.

The Business Research Company. (2019). *Global Fintech Market Value is Expected to Reach $309.98 Billion at a CAGR Of 24.8% Through 2022* [Online; Accessed 01.04.2021]. %5Chttps://www.prnewswire.com/news-releases/global-fintech-market-value-is-expected-to-reach-309-98-billion-at-a-cagr-of-24-8-through-2022--300926069.html

Vaishnavi, V. K., & Kuechler, W. (2015). *Design science research methods and patterns: innovating information and communication technology*. Crc Press.

Vaughan-Nichols, S. (2013). *Apple iPhone fingerprint reader confirmed as easy to hack* [Online; Accessed: 01.10.2020]. http://zdnet.com/apple-iphone-fingerprint-reader-confirmed-as-easy-to-hack-7000021065

Vignolo, J. (2019). *2019 Gartner Market Guide for Online Fraud Detection* [Online; Accessed 12.02.2020]. https://blog.sift.com/2019/gartner-fraud-detection-guide-5-key-takeaways/

Visa. (2016). *Annual Digital Payments Study Poland 2016* [Online; Accessed 02.02.2020]. https://resources.mynewsdesk.com/image/upload/thv1p2ep6thuchr66z6m.pdf

Visa. (2017a). *Goodbye, passwords. Hello, biometrics - Visa biometrics study* [Online; Accessed 07.12.2019]. https://usa.visa.com/dam/VCOM/global/visa-everywhere/documents/visa-biometrics-payments-study.pdf

Visa. (2017b). *Visa - Annual Digital Payments Study 2017* [Online; Accessed 07.12.2018]. https://www.visaeurope.com/newsroom/news/mobile-money-takes-off-as-77-of-europeans-use-their-phones-to-bank-and-make-everyday-payments

Visa. (2017c). *Visa - Digital Payments Study 2017* [Online; Accessed 07.12.2018]. https://www.visa.pl/o-nas/aktualnosci/upowszechnienie-pieniadza-mobilnego-w-polsce-77-percent-badanych-uzywa-smartfonow-do-bankowania-i-codziennych-platnosci-2190949

Voris, J. (2018). Measuring How We Play: Authenticating Users with Touchscreen Gameplay. *International Conference on Mobile Computing, Applications, and Services*, 144–164.

Wang, Q., Su, X., Cai, Z., & Zhang, X. (2017). Mobile iris recognition via fusing different kinds of features. *Chinese Conference on Biometric Recognition*, 401–410.

Weisbaum, H. (2014). *Most americans don't secure their smartphones* [Accessed: 2020-06-09]. http://www.cnbc.com/2014/04/26/most-americans-dont-secure-their-smartphones.html

Wójtowicz, A., & Joachimiak, K. (2016). Model for adaptable context-based biometric authentication for mobile devices. *Personal and Ubiquitous Computing*, *20*(2), 195–207.

Wu, C., He, K., Chen, J., Zhao, Z., & Du, R. (2020). Liveness is Not Enough: Enhancing Fingerprint Authentication with Behavioral Biometrics to Defeat Puppet Attacks. *29th {USENIX} Security Symposium ({USENIX} Security 20)*, 2219–2236.

www.androidauthority.com. (2011). *Galaxy Nexus Face Unlock Works with Real Face AND Photo* [Online; Accessed: 19.07.2020]. http://www.androidauthority.com/galaxy-nexus-face-unlock-works-with-real-face-and-photo-32210/

www.blik.com. (2020). *Almost 100 million transactions and 5.5 million active users* [Online; Accessed 01.04.2021]. https://blik.com/en/blisko-100-mln-transakcji-i-5-5-mln-aktywnych-uzytkownikow

www.csid.com. (2012). *Consumer Survery: Password Habits* [Accessed: 2016-08-09].

www.dotpay.pl. (2020). *Apple Pay w Polsce (Apple Pay in Poland)* [Online; Accessed 12.07.2020]. https://www.dotpay.pl/blog/metody-platnosci/apple-pay-nowoczesna-i-bezpieczna-metoda-platnosci-mobilnych

www.itweb.co.za. (2018). *Smartphone fingerprint scanning heads to mainstream* [Accessed: 2020-06-09]. https://www.itweb.co.za/content/KPNG878XKxX74mwD

www.mathworks.com. (2020). *Android acceleration and gyroscope axes explanation* [Online; Accessed 12.07.2020]. https://www.mathworks.com/help/supportpkg/android/

www.mcafee.com. (2013). *More Than 30% of People Don't Password Protect Their Mobile Devices* [Accessed: 2020-08-09]. https://blogs.mcafee.com/consumer/unprotected-mobile-devices/

www.sophos.com. (2011). *67 Percent of Consumers Don't Have Password Protection on Their Mobile Phones* [Accessed: 2021-02-02]. https://www.sophos.com/en-us/press-

office/press-releases/2011/08/67-percent-of-consumers-do-not-have-password-protection-on-their-mobile-phones.aspx

wwww.en.profit.me. (2017). *Examples of android gestures* [Online; Accessed 10.10.2019]. https://en.proft.me/media/android/android_gestures.jpg

www.worldbank.org. (2020). *Fintech Market Reports Rapid Growth During COVID-19 Pandemic*. https://www.worldbank.org/en/news/press-release/2020/12/03/fintech-market-reports-rapid-growth-during-covid-19-pandemic

Xu, H., Zhou, Y., & Lyu, M. R. (2014). Towards continuous and passive authentication via touch biometrics: An experimental study on smartphones. *Symposium On Usable Privacy and Security, SOUPS*, *14*, 187–198.

Yampolskiy, R. V., & Govindaraju, V. (2008). Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, *1*(1), 81–113.

Yan, J. J., Blackwell, A. F., Anderson, R. J., & Grant, A. (2004). Password Memorability and Security: Empirical Results. *IEEE Security & privacy*, *2*(5), 25–31.

Yang, Y., Guo, B., Wang, Z., Li, M., Yu, Z., & Zhou, X. (2019). BehaveSense: Continuous authentication for security-sensitive mobile apps using behavioral biometrics. *Ad Hoc Networks*, *84*, 9–18.

Zakonnik, Ł., & Czerwonka, P. (2014). PŁATNOŚCI MOBILNE W POLSCE–ANALIZA SWOT. *Studia i Materialy Polskiego Stowarzyszenia Zarzadzania Wiedza/Studies & Proceedings Polish Association for Knowledge Management*, (71).

Zhang, H., Patel, V. M., Fathy, M., & Chellappa, R. (2015). Touch gesture-based active user authentication using dictionaries. *2015 IEEE Winter Conference on Applications of Computer Vision*, 207–214.

Zhao, X., Feng, T., & Shi, W. (2013). Continuous mobile authentication using a novel graphic touch gesture feature. *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 1–6.

Zou, B., & Li, Y. (2018). Touch-based smartphone authentication using import vector domain description. *2018 IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, 1–4.

Zou, L., He, Q., & Feng, X. (2015). Cell phone verification from speech recordings using sparse representation. *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 1787–1791.

Związek Banków Polskich. (2009). Forum Technologii Bankowych. *Biometria w bankowości i administracji publicznej, Warszawa*.

List of Tables

List of Figures

Listings