




# Mobile Device Background Sensors: Authentication vs Privacy

Paula Delgado de Santos

A Thesis submitted for the degree of:
*Doctor of Philosophy in Electronic Engineering*, University of Kent
*Doctor of Philosophy in Computer and Telecommunication Engineering*, UAM

School of Engineering, University of Kent
Escuela Politécnica Superior, Universidad Autónoma de Madrid

July 2023

Departments: School of Engineering
University of Kent, UK

Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid, SPAIN

PhD Thesis: Mobile Device Background Sensors: Authentication vs Privacy

Author: **Paula Delgado de Santos**
Telecommunication Engineer
(Universidad Autónoma de Madrid, SPAIN)

Advisors: **Rubén Tolosana Moranchel**
Doctor in Computer and Telecommunication Engineering
(Universidad Autónoma de Madrid, SPAIN)
**Richard Guest**
Doctor in Electronic Engineering
(University of Kent, UK)

Year: 2023

The research described in this Thesis was carried out within the School of Engineering, University of Kent and the Biometrics and Data Pattern Analytics Laboratory - BiDA Lab at the Dept. of Tecnología Electrónica y de las Comunicaciones, Escuela Politécnica Superior, Universidad Autónoma de Madrid (from 2020 to 2023). This PhD Thesis has been funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860315.

# Abstract

The increasing number of mobile devices in recent years has caused the collection of a large amount of personal information that needs to be protected. To this aim, behavioural biometrics has become very popular. But, what is the discriminative power of mobile behavioural biometrics in real scenarios?

With the success of Deep Learning (DL), architectures based on Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), such as Long Short-Term Memory (LSTM), have shown improvements compared to traditional machine learning methods. However, these DL architectures still have limitations that need to be addressed. In response, new DL architectures like Transformers have emerged. The question is, can these new Transformers outperform previous biometric approaches?

To answers to these questions, this thesis focuses on behavioural biometric authentication with data acquired from mobile background sensors (i.e., accelerometers and gyroscopes). In addition, to the best of our knowledge, this is the first thesis that explores and proposes novel behavioural biometric systems based on Transformers, achieving state-of-the-art results in gait, swipe, and keystroke biometrics.

The adoption of biometrics requires a balance between security and privacy. Biometric modalities provide a unique and inherently personal approach for authentication. Nevertheless, biometrics also give rise to concerns regarding the invasion of personal privacy. According to the General Data Protection Regulation (GDPR) introduced by the European Union, personal data such as biometric data are sensitive and must be used and protected properly. This thesis analyses the impact of sensitive data in the performance of biometric systems and proposes a novel unsupervised privacy-preserving approach.

The research conducted in this thesis makes significant contributions, including: *i)* a comprehensive review of the privacy vulnerabilities of mobile device sensors, covering metrics for quantifying privacy in relation to sensitive data, along with protection methods for safeguarding sensitive information; *ii)* an analysis of authentication systems for behavioural biometrics on mobile devices (i.e., gait, swipe, and keystroke), being the first thesis that explores the potential of Transformers for behavioural biometrics, introducing novel architectures that outperform the state of the art; and *iii)* a novel privacy-preserving approach for mobile biometric gait verification using unsupervised learning techniques, ensuring the protection of sensitive data during the verification process.

*Don't cry because it's over,*
*smile because it happened*

Dr. Seuss

# Acknowledgements

Firstly, I would like to thank my co-supervisor at UAM, Ruben Tolosana, for all the help during this thesis and the previous years we have worked together. He has advised and guided me at all times, getting me to where I am today. I know that it would not have been possible with another person by my side. I will be eternally grateful.

Also, I also wanted to thank my other co-supervisor at Kent, Richard Guest, for all the support and guidance throughout the PhD. These years have helped me a lot to grow as a researcher.

Following with Kent, I would love to thank all my colleges for all the times in the lab, lunch, coffees, and all the times we have complained together. Thank you for being in the same moment as me, and understanding what I was going through. I would especially like to thank Chantal, Rania, and Pavlos, I hope we will all meet again soon!

Also, to all the people of my Marie Curie and sister networks, PriMa and TReSPAsS, for these years that allowed us to travel a bit (even though at the beginning it was online...) and to get to know each other a bit better! I hope we will continue to grow together.

Lastly in English (o español) to my little family in Canterbury, "Feria de Canterbury". Guille, Antonio, Sam, Carla... you are amazing! I'm taking the best with me! And we will always be connected! Angela and Cris... I think you know that you are two pillars in my life and that won't change! There were no better people to live all this, ¡os quiero mucho mucho mucho! Sea donde sea, sabemos donde encontrarnos.

Time to Spanish...

También quiero agradecer a toda la gente del BiDA Lab, que, aunque mayormente en la distancia, me han ayudado mucho y han sido muy participes de mi doctorado. En especial, a Giuseppe, que hemos sabido como colaborar juntos y me llevo muchos momentos buenos como compañero y como amigo.

No podían faltar todos mis amigos de Segovia, Ayllón, teleco... saber que por muy

lejos que me vaya siempre vais a estar ahí me da fuerzas para seguir. Gracias por apoyarme en esta locura del doctorado y por secar mis lágrimas cuando era necesario. Gracias por no cambiar y seguir teniendo mil momentos juntos. Gracias por enseñarme a cuidar la amistad, y saber lo importante de su significado.

Por último, quiero agradecer a mi familia. Gracias a todos por confiar siempre en mi y por estar orgullosos de mí. Me hace muy feliz saber que os alegrais de todos mis logros igual o más que yo. Tengo a la mejor familia del mundo. En particular, papá, mamá y Sergio, gracias por aguantar esas videollamadas eternas quejándome de todo, vuestros consejos y siempre apoyarme. Gracias por ir conmigo al fin del mundo si es necesario. ¡Os quiero muchísimo!

Paula Delgado de Santos

Canterbury, July 2023

# Glossary

- **1D**: One-Dimension.
- **2D**: Two-Dimensions.
- **5G**: Fifth-Generation.
- **6G**: Sixth-Generation.
- **A-C**: Auto-Correlation.
- **ADE**: Adversary's Estimate.
- **ADR**: Adversary's Resource.
- **AE**: Autoencoder.
- **AER**: Average Error Rate.
- **ANN**: Artificial Neural Network.
- **AR**: Augmented Reality.
- **AUC**: Area Under the Receiver Operating Characteristic.
- **AUD**: Active User Detection.
- **BD**: Bipolar Disorder.
- **BMI**: Body Mass Index.
- **BPM**: Beats Per Minute.
- **BPET**: Biometric Privacy-Enhancing Technologies.
- **B-SVM**: Binary Support Vector Machine.
- **CA**: Continuous Autentication.
- **CMC**: Cumulative Match Characteristic.
- **CNN**: Convolutional Neural Network.

- **COCR**: Correct Overall Classification Rate.
- **CSI**: Channel State Information.
- **DBSCAN**: Density-Based Spatial Clustering of Applications.
- **DET**: Detection Error Trade-off Curve.
- **DL**: Deep Learning.
- **DT**: Decision Tree.
- **DTW**: Dynamic Time Warping.
- **ECG**: Electrocardiograph.
- **EER**: Equal Error Rate.
- **EM**: Expectation Maximisation Clustering.
- **FAR**: False Acceptance Rate.
- **FFT**: Fast Fourier Transform.
- **FRR**: False Rejection Rate.
- **GAN**: Generative Adversarial Network.
- **GBR**: Gait Biometric Recognition.
- **GDI**: Gait Dynamic Images.
- **GDPR**: General Data Protection Regulation.
- **GMM**: Gaussian Mixture Model.
- **GPS**: Global Positioning System.
- **GPU**: Graphics Processing Unit.
- **GRE**: Gaussian Range Encoding.
- **HAR**: Human Activity Recognition.
- **HF**: Hidden Failure.
- **HMM**: Hidden Markov Models.
- **HOS**: Higher-Order Statistics.
- **IF**: Isolation Forest.
- **IMU**: Inertial Measurement Unit.
- **ITN**: Innovative Training Networks.
- **IVE**: Incremental Variable Eliminator.

- **KDE**: Kernel Density Estimation.
- **KL-Score**: Kullback-Leibler Score.
- $k$**NN**: $k$-Nearest Neighbours.
- **LFW**: Labeled Faces in the Wild.
- **LR**: Logistic Regression.
- **LSTF**: Long Sequence Time-series Forecasting.
- **LSTM**: Long Short-Term Memory.
- **ML**: Machine Learning.
- **NV**: Naïve Bayes.
- **OC-SVM**: One-Class Support Vector Machine.
- **OSA**: Obstructive Sleep Apnea.
- **PCA**: Principal Component Analysis.
- **PDFs**: Probability Density Functions.
- **PFRNet**: Privacy-Enhancing Face-Representation learning Network.
- **PII**: Personal Identifiable Information.
- **PIN**: Personal Identification Number.
- **PPDM**: Privacy Preserving Data Mining.
- **PriMa**: Privacy Matters.
- **PSO**: Particle Swarm Optimisation.
- **RBFN**: Radial Basis Function Network.
- **ReLU**: Rectified Linear Unit.
- **RF**: Random Forest.
- **RL**: Recurrent Layer.
- **RNN**: Recurrent Neural Network.
- **ROC**: Receiver Operating Characteristic.
- **RSSI**: Receiver Signal Strength Indicator.
- **SAN**: Semi-Adversarial Network.
- **SGD**: Stochastic Gradient Descent.
- **SMC**: Secure Multiparty Computation.

- **SpO2**: Saturation of Peripheral Oxygen.
- **SRBD**: Sleep-Related Breathing Disorders.
- **SSID**: Service Set Identifier.
- **SVM**: Support Vector Machine.
- **TASR**: Task Assignment Success Rate.
- **TDA**: Time Delay Aggregation.
- **THAT**: Two-stream Convolution Augmented Human Activity Transformer.
- **TO**: True Outcome.
- **TReSPAsS**: TRaining in Secure and PrivAcy-preserving biometricS.

# Table of Contents

# List of Figures

# List of Tables

# Part I

# Problem Statement and Contributions

# Chapter 1

# Introduction

Why is it necessary to protect mobile devices? What is the safest and least invasive way to protect them? The rapid and continuous deployment of mobile devices in our society has become ubiquitous, interacting with them anytime, anywhere. As a result, our mobile devices have become data hubs, storing all our personal information such as diary and financial information (Delgado-Santos et al., 2022a; Niknejad et al., 2020). As a consequence, it is crucial to protect the access to them using robust and user-friendly techniques ensuring high security at the same time (Melzi et al., 2022a).

Passwords have conventionally served as a means to enhance security, owing to their capacity for safeguarding sensitive information. Although these methods have been widely used and accepted by the society over the years, they have several disadvantages such as being easily stolen or lost. In order to address these limitations, alternative authentication approaches based on biometrics have emerged, offering enhanced protection for devices in a more secure, efficient, and user-friendly way (Jain et al., 2007, 2016).

Biometrics is a scientific field that focuses on identifying individuals based on their

physical or behavioural characteristics, such as their facial features, fingerprints, or handwriting. By harnessing these unique traits, biometric technology offers the potential to integrate into extensive authentication systems and deliver promising outcomes (Jain et al., 2007). In particular, behavioural traits such as gait (Hadjkacem et al., 2020), keystroke dynamics (Stragapede et al., 2022a), touch gestures (Acien et al., 2021a), or handwritten signature (Tolosana et al., 2022b) have recently shown impressive results in different security scenarios, such as subject authentication on mobile devices.

In the field of biometric recognition systems, the traditional approach has relied upon handcrafted features extracted for specific tasks. However, a notable shift in this paradigm has emerged in recent years. This change can be attributed primarily to the abundance of available data and the increased computational resources at our disposal. Consequently, Deep Learning (DL) has emerged, enabling computers to learn from experience and comprehend the world through a hierarchical understanding of simpler components (Minaee et al., 2023).

This thesis is mainly focused on behavioural biometric authentication systems based on DL, such as gait and swipe gestures. In particular, we use data acquired using background sensors (i.e., accelerometer and gyroscope) from mobile devices. Moreover, personal privacy is also an important issue to address when it comes to securing mobile devices. For this reason, this thesis also analyses the impact of privacy in the authentication performance, presenting novel unsupervised privacy-preserving approaches.

This introductory Chapter first presents in Section 1.1 the key aspects of mobile devices nowadays. Then, the deployment of biometric technologies across these devices is presented in Section 1.2, with emphasis on the different modalities and applications of these systems and mobile behavioural traits. Section 1.3 explains the main concepts of DL models. Later on we motivate in Section 1.4 the key aspects of personal data

as defined in the General Data Protection Regulation (GDPR) (GDP, 2016), including the importance of privacy preservation for biometrics. We finish this Chapter with the motivation of the thesis in Section 1.5, the outline of the thesis in Section 1.7 and the detailed research contributions in Section 1.8.

## 1.1 Mobile Devices

Mobile devices, including smartphones, tablets, and smartwatches, have emerged as compact electronic devices designed to offer exceptional portability. These devices have undergone a rapid evolution, characterised by substantial advances in storage capacity and computing capabilities, establishing themselves as primary platforms for communication and interaction.

Furthermore, modern mobile devices are equipped with several sensors that greatly improve our daily lives, finding utility in diverse domains such as lifestyle computing, medicine, sports, and personal security (Niknejad et al., 2020; Wright and Keith, 2014). These sensors enable a wide range of functionalities, providing people with features that enhance their experiences and facilitate various tasks. The continuous and prolonged use of mobile devices over time has led to a huge collection of data, converting the devices into small datahubs (Delgado-Santos et al., 2022a).

However, within all these collected data, mobile devices can also acquire a vast amount of personal information. It is therefore important to provide high-security authentication methods, where only authorised individuals (typically the devices owners) has access. To this end, biometrics has emerged as a relevant topic for authentication on mobile devices ensuring data protection and high security at the same time (Abuhamad et al., 2021; Ellavarason et al., 2020; Jain et al., 2016; Tolosana et al., 2022a).

## 1.2 Biometrics

Biometric traits are inherently unique to an individual, exhibiting a strong and relatively permanent link between a person and their biometric features. This enables biometric recognition to be used as a reliable means of identifying individuals (Jain et al., 2016). The possibility of utilising personal anatomical traits for computer based identity verification was first scientifically demonstrated in the 1960s. In (Trauring, 1963), the author examined the minutiae in finger-ridge patterns, providing evidence of the feasibility of these patterns for automatic identity verification. Despite being written over 50 years ago, it is remarkable to see the prescience of this work in anticipating the numerous applications of biometrics in contemporary times (Fierrez et al., 2005; Jain et al., 2004; Jain et al., 2007, 2016; Ross and Jain, 2004b; Villani et al., 2006).

In the field of biometrics, two operational modes are widely studied: *identification* and *verification* (Jain et al., 2016). Identification involves predicting the identity of a subject based on comparing a biometric sample against all available templates in a database (one-to-many match). The identification process results in either identifying one or more subjects whose templates produce the highest similarity with the query sample, or determining that the sample does not match any templates in the database. The identification process can be categorised as either *close-set*, where the system is forced to output an identity, or *open-set*, where the system can recognise the absence of a match. This operational mode has gained significant importance for law enforcement agencies as it aids in identifying possible terrorists or criminals from a potentially extensive watch list, despite being time-consuming and resource-intensive due to the high number of comparisons required. Verification, on the other hand, involves comparing a query biometric sample with the template of the claimed subject (one-to-one match). The goal is to authenticate the subject's identity based on the presented biometric data

and the pre-existing record (e.g., passport control) (Jain et al., 2007).

Over the years, the field of biometrics has expanded significantly, including various modalities such as face (Schroff et al., 2015), iris (Bowyerin and Burge, 2016), fingerprint (Maltoni et al., 2009), palmprint (Svoboda et al., 2016), ear (Chen et al., 2015), keystroke (Maiorana et al., 2011), handwritten signature (Tolosana et al., 2018), touchscreen gestures (Fierrez et al., 2018b), and voice (Ghahabi and Hernando, 2017), among many others. The abundance of available traits and the impressive performance of biometric systems often lead to the questions: Which biometric modality is the most suitable? And which one should be chosen for a security system?

### 1.2.1 Biometric Systems: Modalities and Applications

Biometric modalities can be categorised into two broad classes: *physiological* and *behavioural*. Physiological biometrics encompass modalities that describe inherent physical characteristics of individuals. Examples include facial features (Schroff et al., 2015), fingerprints (Maltoni et al., 2009), iris patterns (Bowyerin and Burge, 2016), hand geometry (Burgues et al., 2010), palmprint (Svoboda et al., 2016), ear (Chen et al., 2015), and retina patterns (Choraś, 2012), among others. On the contrary, behavioural biometrics capture information related to human actions and behaviours. This includes speech patterns (Dong et al., 2018), signature dynamics (Tolosana et al., 2019), handwriting characteristics (Tolosana et al., 2021a), gait (Nguyen et al., 2017), keystroke patterns (Stragapede et al., 2022a), mouse dynamics (Mondal and Bours, 2017), touchscreen gestures (Fierrez et al., 2018b), and other similar traits.

The potential for any human characteristic to serve as a biometric identifier depends on fulfilling several general requirements (Jain et al., 2007):

- **Universality**, refers to the extent to which a biometric trait is present in the global population.
- **Uniqueness or Distinctiveness**, implies that the trait should be unique to each individual or at least sufficiently discriminative to differentiate between subjects.
- **Permanence**, dictates that the biometric trait should have a relatively stable representation over an extended period of time.
- **Collectability**, signifies that the biometric modality can be easily and quantitatively measured.

In addition to these fundamental requirements, there are practical criteria that are also desirable when considering biometric modalities:

- **Performance**, encompasses factors such as efficiency, accuracy, speed, robustness, and resource requirements of implementing a biometric trait in automatic capture and processing of a system.
- **Acceptability**, relates to the willingness of people to utilise a particular biometric trait for authentication purposes, taking into account the conditions and context in which it is used.
- **Circumvention**, reflects the level of difficulty in deceiving or fooling a system using fraudulent methods based on a specific biometric trait.
- **Cost**, refers to the expenses involved in implementing the system in real-world scenarios.
- **Proportionality**, considers the trade-off between the level of privacy individuals are willing to surrender to the system and the services they expect to receive in return.

Ideally, biometric traits should exhibit low *intra-subject variability*, meaning that the biometric information remains relatively stable across multiple measurements of the same individual. Conversely, they should demonstrate high *inter-subject variability*, indicating that the biometric information differs significantly among individuals. Achieving this balance ensures the effectiveness and reliability of a biometric security approach. However, biometric traits often exhibit variations. The capture of physiological samples such as those face are influenced by environmental factors (Chan et al., 2018; Ding and Tao, 2017) like illumination, background scene, occlusion, pose, or makeup, which can introduce intra-subject variability. Behavioural traits, such as voice or handwritten signature, are influenced not only by the physiological model of the individual but also by mood and other contextual factors, resulting in potentially higher intra-subject variability (Gavrilova et al., 2017). Additionally, accuracy in both physiological and behavioural biometric systems can be affected by sensor interoperability, which refers to the compatibility between acquisition devices used during both enrolment and testing phases. Factors such as resolution, sampling frequency, screen size, frame rate, acquisition spectrum, signal-to-noise ratio, and distance between the subject and the camera can all impact system performance (Alonso-Fernandez et al., 2010; Ross and Jain, 2004a).

Hence, the question of "Which biometric trait is the best one?" raised earlier becomes dependent on various factors. The choice of biometric modality depends on the specific application scenario, subject acceptance, risk level, usability, and feasibility, among other considerations. It is unlikely that a single biometric trait will fulfil the requirements of all applications, leading to the development of multi-biometric systems that fuse multiple modalities. In the context of mobile devices, with their diverse range of sensors and capabilities, it is possible to leverage the sensor availability for deployment of behavioural biometrics. By combining and analysing these various modalities, mobile devices can offer enhanced recognition methods that address the specific requirements of different

applications, ensuring both convenience and security for subjects.

### 1.2.2 Mobile Behavioural Biometrics

Sensors typically included in mobile devices, such as accelerometers, gyroscopes, and touchscreens, allow for the capture of various behavioural traits, including gait patterns, swipe gestures, and keystroke dynamics. By leveraging these sensors and advanced algorithms, mobile devices can provide convenient and reliable recognition methods while ensuring subject privacy and security (Das et al., 2018).

Using these behavioural traits, mobile devices can offer advanced recognition methods that go beyond traditional password-based systems. Instead of relying solely on something the subject knows (like a password), behavioural biometrics utilise inherent traits that are difficult to replicate or fake (Gupta and Tripathy, 2023). This approach enhances security and convenience, as subjects can be authenticated based on their natural patterns.

Behavioural biometrics provide a non-intrusive and user-friendly way to verify the identity of individuals, making it suitable for various applications, including device unlocking, mobile payment systems, and access control (Jain et al., 2016). Additionally, these biometric techniques can adapt to the subject's changing behaviour over time, making them more robust and reliable in different scenarios (Shen et al., 2018). However, What is the discriminative power of behavioural biometrics? Why are mobile behavioural biometrics important? What advantages does these traits offer?

## 1.3 Deep Learning

DL has become an increasingly important topic in recent years, empowering computers to learn from experience and comprehend the world by breaking it down into hierarchical components. This innovative approach has led to significant progress in a wide variety of practical applications such as natural language processing, computer vision, biometrics, and many others (Goodfellow et al., 2016; Wang et al., 2020). The field of biometrics has also capitalized on DL, with notable applications in speech recognition (Dong et al., 2018), handwritten signature (Tolosana et al., 2021b), facial recognition (Sun et al., 2020), fingerprint recognition (Darlow and Rosman, 2017), and even Presentation Attack Detection (PAD) (Ramachandra and Busch, 2017) and digital manipulations such as DeepFakes (Tolosana et al., 2020), amongst others.

Two of the most popular techniques are Convolutional Neural Networks (CNNs) (Goodfellow et al., 2016; Gu et al., 2018; Li et al., 2021b) and Recurrent Neural Networks (RNNs) (Medsker and Jain, 2001). Although these networks have been introduced in many deep learning applications, they still have certain limitations. One of the main drawbacks is that they have difficulty capturing long-range dependencies and understanding the context of information that is widely separated. This limitation can affect the performance and accuracy of models, especially when dealing with complex sequences or data with long-range dependencies. To overcome these limitations, new DL architectures such as Transformers have emerged (Hutchins et al., 2022; Vaswani et al., 2017).

Transformers are designed to capture both local and global dependencies in the data, enabling them to understand complex relationships and context over long distances. They achieve this through a novel mechanism which allows them to focus on different

parts of the input data and consider their dependencies when making predictions. These architectures have already garnered immense interest due to their effectiveness across a range of application domains such as language, vision, and reinforcement learning (Tay et al., 2022).

In this context, the following questions come to mind: Can these new Transformer architectures be used for biometrics in order to outperform previous ones? Which Transformer architecture should be proposed? Is it necessary to adapt the Transformer architecture to each specific biometric task?

## 1.4 Privacy Preservation

The adoption of biometrics introduces the necessity to strike a delicate balance between security and privacy. Biometric modalities offer a distinctive and inherently personal method of recognition, ensuring enhanced security measures. However, due to their inherent nature, these traits also raise concerns regarding the potential misuse of personal information. The extensive availability of personal data generated on mobile devices, in combination with device ubiquity (with 6.6 billion mobile devices globally in 2022, estimated to rise to 7.6 billion by 2026 (SMA, 2023)) and their *always-on* nature has turned this technology into a potential source of major invasion of personal privacy.

The European Union has created the GDPR, defining personal data as any information related to an identified or identifiable natural person (GDP, 2016). Moreover, the GDPR also defines sensitive data as a subset of personal information, that includes: *i)* personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; *ii)* trade-union membership; *iii)* genetic data and biometric data processed solely to identify a human being; *iv)* health-related data; and *v)* data concerning a per-

son's sex life or sexual orientation. Automated profiling of subject data (GDP, 2016), can easily reveal such attributes from data acquired through mobile subject interaction by requesting irrelevant permissions, poorly described definitions of permissions, or inappropriate use of permissions combined with the aggregation of highly personalised data, reducing the privacy and security experience of the subjects (Aljeraisy et al., 2021; Barth et al., 2019). Consequently, many works in the literature have focused on preventing potential inappropriate use of data. This is the motivation of recent EU-funded Innovative Training Networks (ITN) such as PriMa (Privacy Matters) (Pri, 2019) and TReSPAsS (TRe, 2019).

The use of sensitive data, such as gender or ethnicity, within biometric modalities has been known to enhance the performance of systems, but it raises concerns regarding subject privacy. In this context, it is crucial to distinguish between privacy protection and sensitive data protection (GDP, 2016). Both aspects aim to de-identify subject data and prevent re-identification (Garfinkel, 2015) of direct identifiers such as names, social security numbers, and addresses, as well as indirect identifiers that, when combined with other information, can potentially identify individuals (ISO, 2017). Privacy protection specifically focuses on securing personal data and draws upon terminology, definitions, and methods from cybersecurity (Agrawal et al., 1993). Moreover, sensitive data protection focuses on techniques for modifying data that account for sensitive information while maximising the utility of the remaining data for analysis (Dalenius, 1986). A pertinent question arises: How different would the biometric performance be if we remove this sensitive information? Is it possible to achieve a balance between privacy and authentication performance?

## 1.5 Motivation of the Thesis

The research conducted in this thesis has been primarily motivated by the following five observations:

The first observation pertains to the significant proliferation of mobile devices, with nearly two-thirds of the world population owning one. It is projected that the number of mobile device users will reach 7.5 billion by 2026 (SMA, 2023). As a result, mobile devices have become an integral part of individuals' lives, being utilised for several daily activities (Salehan and Negahban, 2013). Consequently, ensuring the security of these devices is of utmost importance, considering the vast amount of personal information they collect.

The second observation is a direct consequence of the first. While there are numerous techniques available to secure these devices, it is crucial to select the most suitable approach that caters to the specific needs of the individuals. In this regard, robust and user-friendly techniques, such as biometrics, play a pivotal role in simplifying our daily lives (Melzi et al., 2022a).

The third observation revolves around subject acceptance. The employed techniques must be embraced by individuals (Das et al., 2018). Behavioural biometrics, for instance, has gained wide acceptance amongst the general public due to its non-intrusive nature and difficulty of replication. Individuals do not need to engage in any specific activities; their behavioural data can be extracted simply by utilising their devices as part of their routine (e.g., gait data while walking with the device, swipe data while reading, or keystroke data while typing) (Delac and Grgic, 2004; Jain et al., 2007).

The fourth observation stems from the remarkable accomplishments of DL in vari-

ous domains of behavioural biometrics. However, the use of Transformers, a novel DL technique, remains unexplored until now. Transformers have exhibited remarkable outcomes in domains such as language, vision, and reinforcement learning, surpassing the capabilities of previous DL systems like CNN and RNN (Tay et al., 2022).

The final observation emphasises the importance of not only protecting mobile devices but also preserving privacy. The automated processing of personal data, often referred to as profiling, can potentially expose sensitive attributes by requesting unnecessary permissions during mobile interactions. Therefore, while securing the device is undoubtedly crucial, safeguarding the privacy of the data employed is equally significant (GDP, 2016).

## 1.6 Research Questions

The following are the research questions that will be answered during this thesis:

- Do the mobile background sensors sufficiently support behavioural biometrics? Can these sensors be effectively utilised to identify subjects?
- Can these new Transformer architectures outperform previous behavioural biometric approaches such as CNNs and RNNs?
- Is it possible to achieve a balance between privacy and authentication performance in behavioural biometrics, such as in gait?

It is crucial to note that certain significant aspects are not within the scope of our current investigation, and we acknowledge them as areas for future exploration due to constraints such as time limitations. An example of this is the consideration of active attack scenarios.

## 1.7 Outline of the Thesis

This thesis is organised into four distinct parts as illustrated in Fig. 1.1. *Part 1* establishes the problem statement and outlines the contributions made by this research. Subsequently, two experimental parts are included; *Part II* focuses on the experimental work undertaken in the field of mobile biometric authentication, whilst *Part III* investigates mobile biometric privacy, providing a novel solution. Finally, the thesis concludes with *Part IV*. The structure of the Chapters is as follows:

- *Part I: Problem Statement and Contributions*
  - Chapter 1 provides an introduction to the key aspects addressed in this thesis: mobile devices, biometrics, modalities and applications of biometric systems, mobile behavioural biometrics, deep learning models, and privacy preservation for mobile biometrics.
  - Chapter 2 provides a comprehensive overview of the sensors and raw data commonly found in modern mobile devices, with a particular focus on the background sensors. Additionally, we provide insights into the typical application scenarios and the purposes for which data is collected from mobile devices. Furthermore, we present an extensive review of relevant literature aligned with the themes explored in this thesis, including mobile biometric authentication and privacy. Specifically, we investigate gait and swipe biometrics for recognition purposes. Lastly, we present a summary of metrics proposed in existing literature for quantifying privacy concerns associated with sensitive data, alongside a comprehensive examination of methods employed to protect such data.

*Part II: Mobile Biometric Authentication*

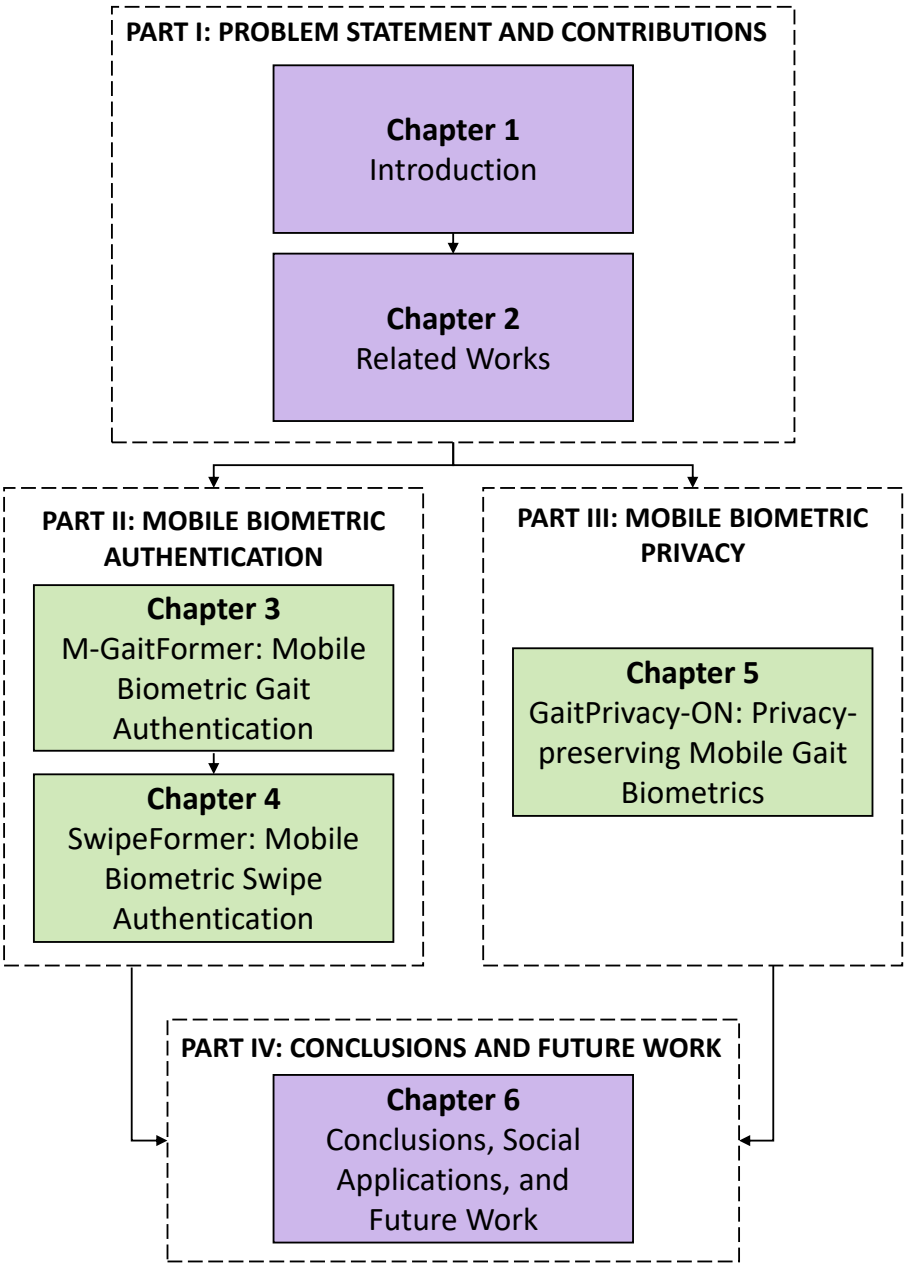



**Figure. 1.1:** Overview of this thesis and dependence among Chapters.

  - Chapter 3 focuses on gait biometric recognition. First we analyse two widely recognised public databases, whuGAIT (Zou et al., 2020) and OU-ISIR databases (Ngo et al., 2014), for the task of recognition with background sensors on mobile devices. Then, we explore and propose novel gait biometric systems based on Transformers, M-GaitFormer. Lastly, an extensive analysis of the proposed M-GaitFormer is presented within two experimental frameworks, catering to both identification and verification purposes.
  - Chapter 4 addresses the swipe biometric verification task. First, we present an in-house database collected in real operational conditions (in-the-wild). In addition, examines two popular publicly available databases collected under constrained conditions: Frank DB (Frank et al., 2012) and HuMIdb (Acien et al., 2021a). Then, we propose SwipeFormer, a novel touchscreen biometric verification system based on Transformers. A comprehensive analysis of the proposed SwipeFormer system is then conducted, with a particular focus on the verification task. Additionally, we include a case study involving the application of these Transformer-based systems to another behavioural biometric trait, specifically addressing the challenging free-text keystroke mobile scenario. To facilitate a comprehensive evaluation, we carry out experimental frameworks encompassing the proposed systems.

- *Part III: Mobile Biometric Privacy*

  - Chapter 5 presents a novel privacy-preserving mobile gait verification approach based on unsupervised learning, GaitPrivacyON. In addition, we include an in-depth quantitative analysis of GaitPrivacyON over three popular databases in the field of gait recognition, MotionSense (Malekzadeh et al., 2018), MobiAct (Vavoulas et al., 2016), and OU-ISIR (Ngo et al., 2014).

- *Part IV: Conclusions, Social Applications, and Future Work*

– Chapter 6 concludes the thesis summarising the main results obtained, social applications of the thesis, and provides new research lines to continue the work carried out during this thesis.

Regarding the dependence among Chapters, before reading any of the experimental Chapters 3, 4, and 5 (green boxes in Fig. 1.1), it is recommended to read the Chapters related to problem statement and contributions.

## 1.8 Detailed Research Contributions

The research contributions of this thesis are the following (some publications appear in several items of the list):

- PART I: PROBLEM STATEMENT AND CONTRIBUTIONS:

  1. State of the Art

     – **P. Delgado-Santos**, G. Stragapede, R. Tolosana, R. Guest, F. Deravi, and R. Vera-Rodriguez, "A Survey of Privacy Vulnerabilities of Mobile Device Sensors", *ACM Computing Surveys*, 54(11), 2022.

     – **P. Delgado-Santos**, R. Tolosana, R. Guest, F. Deravi, and R. Vera-Rodriguez, "Exploring Transformers for Behavioural Biometrics: A Case Study in Gait Recognition", *Pattern Recognition*, 2023.

     – **P. Delgado-Santos**, R. Tolosana, R. Guest, R. Vera-Rodriguez, and J. Fierrez, "M-GaitFormer: Mobile Biometric Gait Verification using Transformers", *Engineering Applications of Artificial Intelligence*, 2023.

     – **P. Delgado-Santos**, R. Tolosana, R. Guest, P. Lamb, A. Khmelnitsky, C. Coughlan, and J. Fierrez, "SwipeFormer: Transformers for Mobile

Touchscreen Biometrics", *Under Review in Expert Systems with Applications*, 2023.

- **P. Delgado-Santos**, R. Tolosana, R. Guest, R. Vera-Rodriguez, F. Deravi, and A. Morales, "GaitPrivacyON: Privacy-Preserving Mobile Gait Biometrics using Unsupervised Learning", *Pattern Recognition Letters*, 161:30–37, 2022.

- PART II: MOBILE BIOMETRIC AUTHENTICATION

  1. Mobile Biometric Gait Authentication

     - **P. Delgado-Santos**, R. Tolosana, R. Guest, F. Deravi, and R. Vera-Rodriguez, "Exploring Transformers for Behavioural Biometrics: A Case Study in Gait Recognition", *Pattern Recognition*, 2023.
     - **P. Delgado-Santos**, R. Tolosana, R. Guest, R. Vera-Rodriguez, and J. Fierrez, "M-GaitFormer: Mobile Biometric Gait Verification using Transformers", *Engineering Applications of Artificial Intelligence*, 2023.

  2. Mobile Biometric Swipe Authentication

     - **P. Delgado-Santos**, R. Tolosana, R. Guest, P. Lamb, A. Khmelnitsky, C. Coughlan, and J. Fierrez, "SwipeFormer: Transformers for Mobile Touchscreen Biometrics", *Under Review in Expert Systems with Applications*, 2023.
     - G. Stragapede, **P. Delgado-Santos**, R. Tolosana, R. Vera-Rodriguez, R. Guest, and A. Morales, "Mobile Keystroke Biometrics using Transformers", *Proc. International Conference on Automatic Face and Gesture Recognition*, 2023.
     - G. Stragapede, **P. Delgado-Santos**, R. Tolosana, R. Vera-Rodriguez, R. Guest, and A. Morales, "TypeFormer: Transformers for Mobile Keystroke Biometrics", *ACM Transactions Computer-Human Interaction*, 2023.

- PART III: MOBILE BIOMETRIC PRIVACY

    1. Privacy-Preserving Mobile Behavioural Biometrics System

        – **P. Delgado-Santos**, R. Tolosana, R. Guest, R. Vera-Rodriguez, F. Deravi, and A. Morales, "GaitPrivacyON: Privacy-Preserving Mobile Gait Biometrics using Unsupervised Learning", *Pattern Recognition Letters*, 161:30–37, 2022.

# Chapter 2

# Related Works

This Chapter summarises previous studies related to this thesis. First, Section 2.1 explains the different background sensors available in mobile devices whereas in Section 2.2 we describe popular mobile application scenarios. Then, Section 2.3 provides an overview of biometric authentication or recognition on mobile devices. In particular, we present in Sections 2.3.1 and 2.3.2 the gait and swipe biometric recognition scenarios, respectively. Section 2.4 provides a description of the state of the art in privacy sensitive data extraction, Finally, we describe in Sections 2.5 and 2.6 the privacy metrics and privacy-preserving techniques presented for mobile biometrics, respectively.

This Chapter is based on the following publications: (Delgado-Santos et al., 2022a,b, 2023a,b,c).

## 2.1 Mobile Background Sensors

Mobile devices offer a rich ground for data collection and processing. Smartphones, to begin with, are distinguished from previous generation cellular phones by their powerful hardware capabilities (e.g., equipped with multi-core processors, GPUs, hardware acceleration units, and gigabytes of memory) and mobile operating systems, which facilitate wider sensing and multimedia software applications (Lai et al., 2017).

Mobile device built-in sensors, known as background sensors, are capable of providing frequent measures of physical quantities in an unobtrusive and transparent way. However, these data can be easily exploited to extract sensitive information of the subject such as gender, age, emotion, ethnic group, etc (Delgado-Santos et al., 2022a).

Similar to background sensors, wearable devices such as smartwatches also possess these capabilities. Wearable devices refer to a broad category of electronic devices that can be worn on the body as accessories or incorporated into clothing or personal items, designed to enhance and augment various aspects of human activities (Wright and Keith, 2014). Their popularity amongst consumer electronics is rapidly increasing and they are progressively becoming capable of more specialised measurements and analyses (John Dian et al., 2020). In general, wearables typically have mobile applications that are installed on smartphones for communication and computing purposes, along with a more comprehensive user interface. For example, smartwatches or fitness tracker bracelets can provide physical measurements such as walking distances using motion sensors and Global Positioning System (GPS). In addition, other interesting information related to physiological parameters such as heart rate, electrocardiogram (ECG), stress, sleep quality, etc. are usually available (Romero-Tapiador et al., 2023).

Table 2.1 provides a description of the sensors and the raw data commonly available

**Table 2.1:** Description of the sensors and the raw data commonly available in modern mobile devices.

| Sensor Type | Sensor / Data Source | Measured / Logged Quantity | Scope / Purpose | Sensor Type |
|---|---|---|---|---|
| Motion | Accelerometer / Linear Accelerometer | Acceleration Force | Device Traslation | Hardware |
| | Gyroscope | Angular Velocity | Device Rotation | Hardware |
| | Rotation Vector | Angle | Device Orientation | Hardware, Software |
| | Gravity | Magnitude of Gravity | Device Orientation | Hardware, Software |
| | Significant Motion | Change of subject movement | Walking or Riding Vehicle | Software |
| | Step Counter | Number of Steps | Physical Activity Tracking | Software |
| | Step Detector | Step | Physical Activity Tracking | Software |
| Position | Geomagnetic Field | Earth's Magnetic Field | Device Orientation | Hardware |
| | Proximity | Distance | Device Distance from Surface | Hardware |
| | Magnetometer | Earth's Magnetic Field | Device Orientation | Hardware |
| | Geomagnetic Rotation Vector | Earth's Magnetic Field | Device Orientation | Hardware, Software |
| | Game Rotation Vector | Angle | Device Rotation | Hardware, Software |
| Environmental | Light | Illuminance | Screen Luminosity Regulation | Hardware |
| | Pressure | Ambient Pressure | Contextual Information | Hardware |
| | Temperature | Ambient Temperature | Contextual Information | Hardware |
| | Humidity | Ambient Humidity | Contextual information | Hardware |
| Health | BPM | Number of Beats | Physical Activity Monitoring | Hardware |
| | ECG | Sinus Rhythm Graph | Physical Activity Monitoring | Hardware |
| | $SpO_2$ | Arterial Blood Oxygen Saturation Percentage Level | Physical Activity Monitoring | Hardware |
| | Blood Pressure | Systolic and Diastolic Average Pressure | Physical Activity Monitoring | Software |
| | Stress | Percentage based on Heart Beat Variability | Physical Activity Monitoring | Software |
| | Sleep / Wake Amount | Time | Physical Activity Monitoring | Hardware, Software |
| | Sleep Phase Transitions | Time | Physical Activity Monitoring | Hardware, Software |
| | Caloric Consumption | Step Counter | Physical Activity Monitoring | Software |
| Touchscreen | Keystroke | Keys Presses and Releases | Key Input | Hardware |
| | Touch Data | Screen Coordinates, Pressure of Touch | Complex Touch Gestures | Hardware |
| Network, Location and Application | Wi-Fi | SSID, RSSI, Encryption Protocol, Frequency, Channel | Connectivity | Hardware |
| | Bluetooth | SSID, RSSI, Encryption Protocol, Frequency, Channel | Connectivity | Hardware |
| | Cell Tower | ID | Connectivity | Hardware |
| | GPS | Latitude, Longitude, Altitude, Bearing, Accuracy | Navigation | Hardware |
| | App Usage | Name and Time of Used Apps | System Log | Software |

BPM- Beats Per Minute, ECG- Electrocardiogram, $SpO_2$- Saturation of Peripheral Oxygen, GPS- Global Positioning System, SSID- Service Set IDentifier, RSSI- Receiver Signal Strength Indicator.

in modern mobile devices, grouped according to their sensing domain. In general, sensors can be classified into two categories based on the process adopted to produce the output signal: *i)* hardware sensors, are physically installed components that perform

a transduction of the physical quantity they measure to an electrical signal, which is converted into the digital domain for further processing; and *ii)* software sensors, rely on data already made available by hardware sensor and/or calculate them to produce a measurement.

Motion sensors are responsible for measuring the acceleration and rotational forces in the three axes of the device. Hardware-based motion sensors register continuous quantities as in the case of acceleration or angular velocity, whereas for the software-based sensors, their output could be either continuous or event-driven as in the case of a step detector. Position sensors range from measuring changes in the Earth's magnetic field for orientation in space to proximity sensors, whereas environmental sensors are generally triggered by an event and return a single scalar value measurement. When designed to return continuous measurements, the sampling rate of these sensors can usually reach up to 200 Hz. Also, their power consumption tends to be low (Acien et al., 2021a).

Specific physiological/biological parameter measurements are also available on many mobile devices thanks to dedicated health sensors. For example, most smartphones and smartwatches include built-in optical sensors to capture changes in blood volume in the arteries under the skin, from which heart-related as well as polysomnographic parameters can be obtained (Chen et al., 2013; Tayfur and Afacan, 2019).

Touchscreen data can be in the form of keystrokes acquired from a virtual keyboard (Morales et al., 2016), or in the form of touch data acquired throughout the subject interaction (Tolosana et al., 2020). In the former case, the keys pressed are logged together with pressure and associated timestamps, for each key press and release. From these raw data it is possible to extract more complex features, such as the hold time, inter-press time, inter-release time, etc. (Acien et al., 2021b). In addition to keystrokes,

further data can be collected from the touchscreen sensors (Fierrez et al., 2018b). In fact, it is possible to track the touch position in terms of $X$ and $Y$ coordinates in the screen reference system, but also pressure information and complex multi-touch gestures such as swipe, pinch, tap, and scroll (Tolosana et al., 2022a). Other complex features that can be extracted from touch data are velocity, acceleration, angle, and trajectory (Tolosana et al., 2020).

Connectivity is yet another fundamental aspect of mobile devices. Their utility and ubiquity stem from the vast spectrum of functionalities they support thanks to many installed network protocols. Network connection data retains information about subjects' routine patterns, therefore it can be used for behavioural profiling and sensitive information extraction (Li and Bours, 2018). With the fifth-generation (5G) standard for cellular networks being commercialised and the sixth-generation (6G) in development, a significant improvement is expected in terms of bit rates and latency between machine-to-machine communications, thus increasing the vast spectrum of functionalities already supported by mobile devices (David and Berndt, 2018).

## 2.2 Mobile Applications

In 2008, the two most popular mobile operating systems, Android and iOS, had less than 500 apps available for download. To date, Android device users are able to download over 2.87 million Apps, followed by the Apple app Store with almost 1.96 million apps (Num, 2021). The possible application scenarios are wide ranging. We describe below some of the most popular application scenarios using mobile sensors.

### 2.2.1 Subject Auhtentication

In traditional recognition schemes, the legitimate subject is expected to have knowledge of a secret such as a PIN code or password to gain access (recognition based on "what-you-know"), or an object, such as a card reader and/or token (recognition based on "what-you-have"), whereas recent recognition schemes largely deployed on mobile devices are based on the "what-you-are" paradigm: some traits of the subject are acquired and processed in order to verify the identity (O'Gorman, 2003). With regard to mobile subject recognition, a common approach is based on biometrics (both physiological and behavioural) (Jain et al., 2016), as in the case of entry-point fingerprint or face-based identification. A severe limitation of these processes consists in the fact that once the device is unlocked, as long as it remains active, an intruder would have unlimited time at their disposal. To provide prolonged protection, several studies have investigated and proved the feasibility of continuous recognition schemes for mobile devices based on behavioural biometrics (Patel et al., 2016). In this case, biometric data would be continuously acquired in a passive way throughout normal device usage to constantly verify the subject's identity. Different aspects such as modality, scenarios, and environment, amongst others, can lead to alterations in the performance of mobile biometric systems (Boakes et al., 2019). Often combined, background sensors (Acien et al., 2019b; Wan et al., 2018), touchscreen (Santopietro et al., 2020), and network information (Li and Bours, 2018) are among the most frequent modalities explored to develop behavioural biometric continuous recognition systems.

### 2.2.2 Healthcare and Fitness

Healthcare is a major field of study for mobile applications. The term "mHealth" was coined to indicate a sub-set of eHealth that includes medical and public health practice

supported by mobile devices. Mobile apps might help to improve healthcare delivery processes, and patients could benefit in terms of monitoring and treatment of diseases and chronic conditions, amongst many other healthcare purposes (Nussbaum et al., 2019). Examples of mobile apps include those that provide measurements of postures, report on mental disorders (Gravenhorst et al., 2015), and assess symptoms of conditions such as Parkinson disease, stress, dementia, etc. (Faundez-Zanuy et al., 2020; Majumder and Deen, 2019). Moreover, mobile health apps can be essential in sustaining a healthy lifestyle by monitoring and recommending behaviour corrections. From this perspective, mobile devices such as smartwatches are largely used for fitness tracking. Physical exercise monitoring takes place by acquiring and processing background and GPS sensor data in an explicit and transparent way for the subject (Anjum and Ilyas, 2013; Antar et al., 2019; Khan et al., 2020).

### 2.2.3 Location-based Services

GPS and geolocation data are used by applications to present information related to the environment and the location of subjects for purposes such as targeted advertising, navigation, and recommendations (Haris et al., 2014). These location-aware applications are under the context awareness paradigm (Saha and Mukherjee, 2003). Additionally, besides their native scope of communication, short-range protocols such as Bluetooth and Wi-Fi allow mobile devices to exploit the information of nearby devices for purposes similar to the ones described. This concept can be particularly useful for defining a semantic context of the immediate surroundings, especially in the case of indoor environments. For example, in (Luca and Alberto, 2016), the authors explored the feasibility of creating virtual tours in museums or expositions to deliver information about the items in the proximity of the subjects, who can receive this information on their mobile devices.

### 2.2.4 Other Applications

Traditionally, background sensors contribute to an improvement of subject experience in several ways. For instance, position sensors are useful for recognising the orientation of the device in order to switch from portrait to landscape mode, and vice versa. Light sensor information is used to automatically adjust the screen brightness. The proximity sensor will lock the screen and activate a different speaker when the subject is making a call. Mobile device background sensors are also widely employed for Augmented Reality (AR) applications in several fields, such as education, entertainment, commerce, and navigation, among others (Kim et al., 2017). AR-based apps heavily rely on the information provided by the background sensors to deliver information.

In addition, the sophisticated sensing capabilities of mobile devices, combined with their vast diffusion, have led to the idea of large-scale sensing, known in the literature as *mobile participatory sensing* (Burke et al., 2006). Individuals with sensing and computing devices volunteer to collectively share data to measure and map phenomena of common interest, in a crowd-sourced fashion (Haris et al., 2014). Applications where mobile participatory sensing has been used include the monitoring of noise pollution, litter, and traffic, among others (Melo et al., 2017).

In conclusion, mobile devices can be considered nowadays for multiple applications. Therefore, a thorough and enriching study of their applications is essential to have a more accurate control of the mobile device environment. In particular, in this thesis we pay special attention to the topic of subject recognition.

## 2.3 Mobile Biometric Authentication

Nowadays, scenarios involving mobile devices have gained significant attention. Mobile biometric authentication or recognition has emerged as a crucial area of research (Jain et al., 2016). Within biometrics, behavioural attributes offer a passive, robust, and user-friendly approach for recognition (Delgado-Santos et al., 2022b; Stragapede et al., 2022b). Furthermore, these traits are captured as low-dimensional time-domain signals, enabling fast acquisition and processing.

This Section provides an overview of the two behavioural modalities considered in this thesis: gait and swipe gestures. Section 2.3.1 describes the state of the art in the field of gait biometrics, while Section 2.3.2 focuses on swipe biometrics. By exploring these modalities, we aim to establish a comprehensive understanding of the existing literature and advancements in the respective domains.

### 2.3.1 Mobile Gait Authentication

Gait biometrics has garnered significant attention in recent years, particularly in surveillance scenarios where popular biometric traits such as face and fingerprint pose several disadvantages. Gait recognition leverages on the movement patterns of individuals, focusing on distinctive characteristics like arm swing amplitude, step frequency, and gait length (Wang et al., 2003). With the exponential rise in the number of mobile devices equipped with highly precise sensors, the interest in gait recognition using mobile devices has grown likewise (Marsico and Mecca, 2019).

One of the popular approaches in this domain relies on Inertial Measurement Units (IMUs), such as accelerometers and gyroscopes (Sprager and Juric, 2015). In this Sec-

tion, we study two scenarios of gait recognition: identification (Section 2.3.1.1) and verification (Section 2.3.1.2).

#### 2.3.1.1 Mobile Gait Identification

Mobile gait identification has undergone extensive research and analysis in recent years and has widespread application in forensic investigations. Previous studies have primarily focused on the development of handcrafted methods for feature extraction. Notably, in (Sprager and Juric, 2015), the authors provided a comprehensive summary of the literature in this domain, along with the adoption of Machine Learning (ML) methods for final evaluation.

In this thesis our primary focus lies in exploring the latest DL advancements for mobile gait identification, proposing novel architectures to increase the performance and reliability. Table 2.2 provides a summary of the most relevant methodologies for gait biometric identification on mobile devices based on DL methods. It is important to highlight that all approaches consider the same experimental protocol proposed in (Zou et al., 2020) for two popular public databases in the literature: *i)* whuGAIT (Zou et al., 2020), which comprises accelerometer and gyroscope data acquired from mobile devices, and *ii)* OU-ISIR (Iwama et al., 2012; Ngo et al., 2014), which includes accelerometer and gyroscope data obtained from IMU sensors.

In the past few years, the research community has focused on DL models to improve the robustness of gait identification systems, extracting more discriminative features. As both the spatial and temporal information of the gait pattern is important for the task, DL architectures based on CNN and RNN have been utilised. One of the earliest systems based on DL models using CNNs was created in (Gadaleta and Rossi, 2018). The authors used CNNs for feature extraction and a Support Vector Machine (SVM)

**Table 2.2:** Summary of most relevant state-of-the-art approaches presented in the literature for mobile gait biometric identification based on DL methods for the whuGait and OU-ISIR evaluation datasets (Ngo et al., 2014; Tran and Choi, 2020). The results are shown in terms of Rank-1 accuracy.

| Category | Year | Reference | Description | Performance | Database |
|---|---|---|---|---|---|
| CNN | 2016 | (Gadaleta and Rossi, 2018) | CNN Feature Extractor + SVM Classifier | 92.91% | whuGAIT |
| | | | | 44.29% | OU-ISIR |
| | 2019 | (Delgado-Escaño et al., 2018) | Fusion CNN + Euclidean Distance | 92.89% | whuGAIT |
| | | | | 40.60% | OU-ISIR |
| LSTM | 2020 | (Watanabe and Kimura, 2020) | End-to-End LSTM | 91.88% | whuGAIT |
| | | | | 66.36% | OU-ISIR |
| | 2020 | (Zou et al., 2020) | End-to-End LSTM | 91.88% | whuGAIT |
| | | | | 66.36% | OU-ISIR |
| | 2021 | (Tran et al., 2021) | End-to-End Multi-LSTM | 93.14% | whuGAIT |
| | | | | 78.92% | OU-ISIR |
| CNN + LSTM | 2016 | (Ordóñez and Roggen, 2016) | Cascaded CNN + LSTM | 92.25% | whuGAIT |
| | | | | 37.33% | OU-ISIR |
| | 2020 | (Zou et al., 2020) | 2-Parallel Branches: CNN + LSTM | 93.52% | whuGAIT |
| | 2021 | (Tran et al., 2021) | 2-Parallel Branches: CNN + Multi-LSTM | 94.15% | whuGAIT |
| | | | | 89.79% | OU-ISIR |

CNN- Convolutional Neural Network; LSTM- Long Short-Term Memory; SVM- Support Vector Machine; A-C: Auto-Correlation; RL: Recurrent Layer; GRE: Gaussian Range Encoding.

for the final classification with 0.15% misclassification rates. The score was obtained in less than five walking cycles with their own collected database. Their results proved how DL methods could extract more discriminative features compared with previous ML methods. The same model was evaluated in (Zou et al., 2020) following a predefined experimental protocol, obtaining an accuracy of 92.91% in the whuGAIT database (Zou et al., 2020), and 44.29% accuracy in the OU-ISIR database (Iwama et al., 2012; Ngo et al., 2014). Another approach based on CNNs was presented in (Delgado-Escaño et al., 2018), dividing the data into two branches, according to each sensor (accelerometer and gyroscope). The output of both branches were concatenated to produce a joint feature vector. Cross-validation was used, achieving 95.20% accuracy with the OU-ISIR database using their own experimental protocol. Following the predefined experimental protocol presented in (Zou et al., 2020), results of 92.89% and 44.29% accuracy were achieved in the whuGAIT and OU-ISIR databases, respectively. However, by using only CNNs, the system focuses mainly on spatial characteristics, leaving out the temporal information.

To overcome this drawback, RNNs were proposed extracting temporal features from the time sequences. In (Watanabe and Kimura, 2020) the authors created an end-to-end RNN with a softmax layer. The model was tested with the experimental protocol presented in (Zou et al., 2020), achieving a 91.88% accuracy with whuGAIT database, and 66.36% accuracy with OU-ISIR database. In (Zou et al., 2020) the authors evaluated RNNs over the OU-ISIR database achieving 78.92% accuracy. They also presented the whuGAIT database and proposed a predefined experimental protocol, achieving 93.14% accuracy.

Hybrid approaches have also been proposed in the literature, resulting in a more complex structure, where the CNN extracts spatial features whilst the RNN-LSTM obtains temporal features. In (Ordóñez and Roggen, 2016), the authors presented Deep-

ConvLSTM, which comprises convolutional layers, followed by recurrent and softmax layers. The model obtained 95.8% F1-score for the activity recognition task with the Opportunity database (Chavarriaga et al., 2013). The system was also evaluated for gait identification in (Zou et al., 2020), achieving 92.25% and 37.33% accuracy for the whuGAIT and OU-ISIR databases, respectively. Also, a hybrid approach with two-parallel branches, one CNN and one RNN-LSTM, was presented in (Zou et al., 2020). The extracted features were independent in each branch, obtaining a view of the raw data with both convolutional and recurrent layers. The resulting features from each branch were concatenated and fed into a fully connected layer. The authors achieved 93.52% accuracy on the presented whuGAIT database.

All above methods are based on initially detecting the gait cycle. The input of the DL models is a time interval between two consecutive occurrences of the gait pattern, i.e., putting the same foot on the ground (Marsico and Mecca, 2019). Gait cycle detection is usually a tedious task that can introduce errors due to sensor restrictions (e.g., noise-sensitivity, sensor specification, body placement, etc.). To solve this problem, in (Tran et al., 2021) a new approach using window-based data segmentation was proposed. The authors used a Multi-RNN model considering fixed-length segments as an input, without the need to extract gait cycles. The authors achieved an accuracy of 93.14% for the whuGAIT database, and 78.92% for the OU-ISIR database. In addition, the same authors introduced a hybrid approach, achieving 94.15% and 89.79% accuracy for the whuGAIT and OU-ISIR databases, respectively.

#### 2.3.1.2 Mobile Gait Verification

Whilst previous research has predominantly emphasised gait identification, gait verification has also garnered significant interest due to the large number of real world

applications, such as mobile recognition. In gait verification, the objective is to determine whether a given gait pattern matches the claimed identity.

Although the literature on mobile gait verification is relatively limited compared to gait identification, there have been notable efforts in exploring various methodologies. This thesis aims to contribute to the advancement of mobile gait verification by investigating novel techniques, including ML and DL approaches, and evaluating them through public databases and benchmarks. Table 2.3 provides a summary of the most relevant DL approaches considered in the literature for biometric gait verification on mobile devices.

The authors in (Ngo et al., 2014) implemented a Dynamic Time Warping (DTW) scheme with verification purposes achieving an Equal Error Rate (EER) of 13.50%. In addition the public available OU-ISIR database was presented. However, the experimental protocol considered in that paper might not be very realistic for operational conditions as the same subjects were considered for both training and testing the system. Despite this, other studies in the literature have followed similar experimental protocols, achieving EER values between 5.00% and 10.00% through handcrafted ML techniques, for example: in (Zhong and Deng, 2014) an approach based on Gait Dynamic Images (GDIs) and $i$-vector was presented, in (Sprager and Juric, 2015) the authors proposed a feature extractor based on Higher-Order Statistics (HOS), and in (Subramanian and Sarkar, 2018) the authors introduced an approach based on the Kabsch alignment.

In recent years, researchers have turned to DL techniques to extract more discriminative features. In (Delgado-Escaño et al., 2018), the authors presented an approach based on CNNs. Data were divided into two branches, one for each sensor (accelerometer and gyroscope). CNN features extracted from each branch were concatenated into a common feature vector, and the Euclidean distance was finally computed in order to

**Table 2.3:** Summary of most relevant state-of-the-art approaches presented in the literature for mobile gait biometric verification based on DL methods for the whuGAIT and OU-ISIR evaluation datasets (Ngo et al., 2014; Tran and Choi, 2020). The results are shown in terms of EER. Note that the symbol * indicates those studies that do not use the standard experimental setup considered in the literature.

| Category | Year | Reference | Description | Performance | Database |
|---|---|---|---|---|---|
| Kabsch Alignment | 2019 | (Subramanian and Sarkar, 2018)* | Kabsch Alignmenet Feature Extractor + Tanimoto Similarity | $> 6.00$ | OU-ISIR |
| CNN | 2017 | (Nguyen et al., 2017) | CNN Feature Extractor + SVM Classifier | 10.43 | OU-ISIR |
| | 2020 | (Tran and Choi, 2020) | CNN Feature Extractor + OC-SVM Classifier | 4.49 | OU-ISIR |
| LSTM | 2019 | (Fernandez-Lopez et al., 2019)* | LSTM Feature Extractor + Euclidean Distance | 7.55 | OU-ISIR |
| | 2020 | (Zou et al., 2020) | | 7.50 | whuGAIT |
| | 2021 | (Tran et al., 2021)* | LSTM Feature Extractor + OC-SVM Classifier | 5.82 | whuGAIT |
| | | | | 6.63 | OU-ISIR |
| CNN + LSTM | 2020 | (Zou et al., 2020) | CNN + LSTM Feature Extractor + OC-SVM Classifier | 6.50 | whuGAIT |
| | 2021 | (Tran et al., 2021)* | CNN + LSTM Feature Extractor + OC-SVM Classifier | 4.52 | whuGAIT |
| | | | | 3.36 | OU-ISIR |

CNN- Convolutional Neural Network; LSTM- Long Short-Term Memory; SVM- Support Vector Machine; OC-SVM- One-Class SVM; B-SVM- Binary SVM.

obtain the similarity score between enrolment and test samples. The authors considering a single task instead of different tasks as in the literature. An EER of 1.10% was obtained over the OU-ISIR database, considering the same subjects for training and testing the gait verification systems. A similar approach was also presented in (Tran and Choi, 2020), considering CNNs as the feature extractor and One-Class Support Vector Machine (OC-SVM) for the similarity computation, achieving 4.49% EER in similar experimental protocol conditions.

As described before, previous approaches in the literature tend to use the same subjects to train and test their gait verification systems. However, this scenario may not be realistic representing operational conditions, as the CNN feature extractor needs to be trained every time new subjects are available. Following this observation, in (Nguyen et al., 2017) the authors presented an approach based on CNNs and SVM, considering different sets of subjects for training and testing. They evaluated their proposed approach with the OU-ISIR database achieving an EER of 10.43%, much higher compared with the case of using the same subjects for training and testing. In (Fernandez-Lopez et al., 2019) the authors incorporated RNN-LSTMs to extract features and then compared them with Euclidean distance. A final EER value of 7.55% was obtained.

Apart from the popular OU-ISIR database, in (Zou et al., 2020) the authors presented the whuGAIT database based on mobile data from gyroscope and accelerometer sensors. They also presented a standard experimental protocol considering different subjects for training and testing the systems. In addition, they proposed an approach based on two CNN-branches (one for each sensor) that are concatenated and introduced into a single RNN-branch. A final 6.50% EER was obtained using an OC-SVM classifier for the final similarity computation. Recently, in (Tran et al., 2021) the authors presented a new approach based on a multi-CNN and multi-RNN system. In both databases, different

subjects were used for training and testing the systems, achieving EER values of 4.52% and 3.36% for the whuGAIT and OU-ISIR databases, respectively.

Finally, it is also interesting to remark that in most studies of gait recognition the authors have focused on well-known CNNs and RNNs. However, these DL architectures still have several disadvantages that must be revisited and improved. The main drawbacks are (Hutchins et al., 2022; Vaswani et al., 2017): *i)* sequential computation, not allowing parallelisation within batches; *ii)* compression and summarising of the previous time samples, limiting the past information seen; and *iii)* vanishing gradients during back-propagation; the forget gate in a RNN removes a small portion of the previous state after each sample. To overcome these limitations, new DL architectures such as Transformers are studied in this thesis.

### 2.3.2 Mobile Swipe Verification

Mobile swipe verification systems are receiving a lot of attention nowadays, despite of the fact that other behavioural traits, such as keystroke dynamics or handwritten signature, have traditionally exhibited higher levels of accuracy (Morales et al., 2016; Stragapede et al., 2022c; Tolosana et al., 2022b). The interest for mobile swipe verification stems from the extensive use of touchscreen gestures, such as swiping or tapping, which typically involve brief and normal interactions. However, these gestures present inherent challenges due to their simplicity and the significant variability exhibited by individuals during the sample donation process (Stragapede et al., 2023a).

Swipe gestures are performed continuously in our daily interactions with devices. Therefore, it is essential to establish robust verification methods based on touchscreen biometrics. This is particularly crucial as our daily interactions with mobile devices predominantly rely on simple swipe gestures. The security improvement of mobile devices

requires the deployment of reliable and continuous authentication or recognition techniques adapted to touchscreen-based interactions, thus enhancing the overall security of devices and the trust of individuals (Frank et al., 2012).

Authentication or recognition based on touchscreen biometrics recognises a subject through touch gestures performed on a mobile device screen. Swipe gestures are the most common tasks in touchscreen verification (Frank et al., 2012). Table 2.4 provides a chronological overview of the main touchscreen biometric verification systems in the literature based on swipe gestures, together with their key aspects. One of the main obstacles in this area, apart from the difficulty of the task itself, is the lack of publicly available databases, as each study usually collects its own data (Lamb et al., 2020). In addition, another problem is the heterogeneity of the settings in each study, making a fair comparison very difficult.

Initially, two authentication modalities can be distinguished in this field: continuous and non-continuous. In the first one, continuous authentication, a subject is verified for a period of time while performing gestures on the touchscreen. One of the first studies in the field was (Frank et al., 2012), presenting the public Frank Database with touchscreen data from 4 different android devices and a total of 41 subjects. The authors proposed a system based on the extraction of 30 handcrafted features and One-Class Support Vector Machine (OC-SVM) classifier, achieving performances between 0.00% EER and 4.00% EER with up to 11 swipes per subject. Furthermore, a subject can also be identified in a non-continuous way, where data are collected beforehand and authentication is performed afterwards (Serwadda et al., 2013). In that study the authors also considered touchscreen data, obtaining performances between 10.50% and 17.20% EER with 28 handcrafted features and Logistic Regression.

In addition to the recognition method, the scenario in which the data are acquired

**Table 2.4:** Summary of state-of-the-art approaches presented in the literature for mobile touchscreen biometric verification based on swipe gestures.

| Study | Database (Public) | CA | Scenario (C/U) | Device | N. of subjects | Features | Dimension Feature Vector | Sessions | System | | Authentication Data/Subject | Best Performance [%] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | Feature Extractor | Classifier/ Distance | | |
| (Frank et al., 2012) | ✓ | ✓ | C | HTC Droid Inc.<br>Google Nexus One<br>Google Nexus S<br>Samsung Galaxy S | 41 | T. (x, y, p, t, area) | 30 | 2 ($\geq$ 1 week) | Handcrafted | OC-SVM | 11 swipes | 0.00 - 4.00 (EER) |
| (Serwadda et al., 2013) | ✓ | ✗ | C | Google Nexus S | 191 | T. (x, y, p, t, area) | 28 | 2 | Handcrafted | Logistic Regression | 80 swipes | 10.50 - 17.20 (EER) |
| (Xu et al., 2014) | ✗ | ✓ | C | Samsung Galaxy S2 | 28 | T. (x, y, p, t, area) | 37 | 6 | Handcrafted | B-SVM | 5 swipes (cross-validation) | < 1.00 (EER) |
| (Feng et al., 2014) | ✗ | ✓ | U | Samsung Galaxy S3<br>Samsung Galaxy S4<br>Google Nexus 4 | 23<br>(+ 100 test) | T. (x, y, t) | 6 | 3 | Handcrafted | DTW<br>+ k-NN | 200 swipes | 90.00 (Accuracy) |
| (Bo et al., 2014) | ✗ | ✓ | U | HTC EVO 3D<br>Samsung Galaxy S3 | 10<br>(+ 90 test) | T (x, y, p, t),<br>Acc., Gyr. | 5 | 1 day data | Handcrafted | OC-SVM | 3 swipes | 1 swipe: 23.00 (FAR)<br>12 swipes: 0.00 (FAR) |
| (Saravanan et al., 2014) | ✗ | ✓ | C | Google Nexus 4 Phone<br>Google Nexus 7 Tablet | 10<br>(+ 10 test) | T. (x, y, p, t) | 4 | - | Handcrafted | OC-SVM<br>+ RF | - | Phone: 97.90 (Accuracy)<br>Tablet: 96.80 (Accuracy) |
| (Zaliva et al., 2015) | ✗ | ✗ | C | Samsung Galaxy S4 | 14 | T. (x, y, z, area) | 24 | 15 minutes | Handcrafted | ANN | 5 swipes | 99.96 (F1-Score) |
| (Lu and Liu, 2015) | ✗ | ✓ | C | Personal | 60 | T. (x, y, p, t, area) | 14 | 1 month | Handcrafted | OC-SVM | 100 swipes | 0.03 (FAR)<br>0.05 (FRR) |
| (Zhang et al., 2015) | ✗ | ✗ | C | iPhone 5S | 50 | T. (x, y, p, t, area) | 27 | 3 | Handcrafted | KDTGR | random<br>80 swipes | 11 swipes:<br>2.91 (EER)<br>Frank DB: 0.34 (EER)<br>Serwadda DB: 1.73 (EER) |
| (Antal et al., 2015) | ✓ | ✓ | C | 4 Android devices | 71 | T. (x, y, p, t, area) | 15 | 1 month | Handcrafted | k-NN | 100 swipes | 1 swipe: 65.00 (Accuracy)<br>20 swipes: 100.00 (Accuracy) |
| (Shen et al., 2015) | ✗ | ✓ | C | Samsung Galaxy N7100<br>Samsung Galaxy N9002<br>Huawei Ascend Mate | 71 | T. (x, y, p, t, area) | 22-27 | 3 | Handcrafted | RF | 640 swipes | 11 swipes: 1.80 (EER) |
| (Sitová et al., 2015) | ✓ | ✓ | C | Samsung Galaxy S4 | 100 | T. (x, y, p, t, area)<br>Acc. (x, y, z)<br>Gyr. (x, y, z)<br>Mag. (x, y, z) | 71 | 4 | Handcrafted | OC-SVM | $\geq$ 80 swipes<br>(2 sessions) | 8.50 (EER) |
| (Mahbub et al., 2016) | ✓ | ✓ | U | Google Nexus 5 | 48 | T. (x, y, p, t) | 24 | 2 months | Handcrafted | RF | 70% swipes | 6 swipes: 22.10 (EER) |
| (Sharma and Enbody, 2017) | ✗ | ✓ | C | Google Nexus 7 | 42 | T. (x, y, p, t, area) | 7 | 40 minutes | Handcrafted | B-SVM | random<br>80 swipes | 7.00 (EER) |
| (Wang et al., 2017) | ✗ | ✓ | U | Google Nexus 2<br>Google Nexus 4<br>Google Nexus 7 | 20 | T. (x, y, p, t, area) | 59 | 4 (1 per device) | Handcrafted | B-SVM | 75% swipes | 80.00 (AUC) |
| (Kumar et al., 2016) | ✗ | ✓ | U | Personal | 28 | T. (x, y, p, t, area) | 5 | 4-7 days | Handcrafted | RF | 50% swipes | 99.33 (Accuracy) |
| (Filippov et al., 2018) | - | ✓ | C | - | 20 | T. (x, y, t, area) | 10 | 1 month | Handcrafted | IF | 2000 swipes | 7.50 (FAR)<br>6.40 (FRR) |
| (Siirtola et al., 2018) | - | ✓ | C | Samsung Galaxy S4 | 100 | T. (x, y, p, t, area)<br>Acc. (x, y, z) | 211 | 4 | Handcrafted | EM | 50% swipes | HMOG DB<br>(Read and walk): 7.00 (EER) |
| (Fierrez et al., 2018b) | ✓ | ✗ | C | - | Frank DB: 41<br>Serwadda DB: 191<br>Antal DB: 71<br>UMDAA-02: 48 | All DB:<br>T. (x, y, p, t, area) | 28 | Frank DB: 2<br>Serwadda DB: 2<br>Antal DB: 71<br>UMDAA-02:<br>2 months | Handcrafted | B-SVM<br>+ GMM | 40 swipes | Frank DB: 3.10 (EER)<br>Serwadda DB: 23.30 (EER)<br>Antal DB: 2.60 (EER)<br>UMDAA-02: 3.60 (EER) |
| (Meng et al., 2018b) | ✗ | ✗ | U | Google Nexus 1 | 48 | T. (x, y, t) | 21 | 20 | Handcrafted | ANN | 60% sessions | 2.40 (AER) |
| (Meng et al., 2018a) | ✗ | ✗ | U | Google Nexus 1 | 60 | T. (x, y, p, t, area) | 9 | 30 | Handcrafted | SVM | 67% sessions | 4.70 (AER) |
| (Syed et al., 2019) | ✗ | ✗ | C | Samsung Tab 210"<br>Samsung Tab 27"<br>Samusng S3<br>HTC EVO 4G LTE | 31 | T. (x, y, p, t, area) | 18 | 8 (2-3 weeks) | Handcrafted | RF | 50% swipes | 3.80 (EER) |
| (Acien et al., 2020)<br>HuMIdb | ✓ | ✗ | C | Personal<br>(Android) | 600 | T. (x, y, p) | 64 | $\leq$ 5 ($\geq$ 1 day) | Handcrafted + LSTM | Eucl. Dist. | 70% swipes | 13.00 (EER) |

CA- Continuous Authentication; C- Constrained; U-Unconstrained; T.- Touch, Acc.- Accelerometer; Gyr.- Gyroscope; x- x axis; y- axis; p- pressure; t- timestamp; RNN (LSTM)- Recurrent Neural Network (Long Short-Term Memory); OC-SVM- One-Class Support Vector Machine; B-SVM- Binary SVM; DTW- Dynamic Time Warping; k-NN- k Nearest Neighbours; RF- Random Forest; ANN- Artificial Neural Network; IF- Isolation Forest; EM- Expectation Maximization Clustering; GMM- Gaussian Mixture Model; Eucl. Dist.- Euclidean Distance; Shrunk Cov.- Shrunk Covariance; KDE- Kernel Density Estimation; EER- Equal Error Rate; FAR- False Aceptance Rate; FRR- False Rejection Rate; AUC- Area Under the Receiver Operating Characteristic; AER- Average Error Rate.

is also crucial. Mainly we can distinguish two groups, constrained and unconstrained (a.k.a. in-the-wild) scenarios. In the constrained scenario, the subjects perform a task where data are analysed in a restricted way, i.e., analysis of swipes in a single direction (vertical or horizontal) and/or the orientation of the device (portrait/landscape) (Frank et al., 2012; Serwadda et al., 2013; Xu et al., 2014). On the contrary, in the unconstrained scenario, data are collected whilst subjects use the device freely (Bo et al., 2014; Feng et al., 2014).

In the past few years, the research community has focused on the manual extraction of an optimal set of features from the touchscreen, and their subsequent input into a ML model used as a classifier for the verification task. The most popular classifier was OC-SVM (Frank et al., 2012). The authors in (Saravanan et al., 2014) were able to achieve 97.90% and 96.80% accuracies using a Google Nexus 4 Phone and a Google Nexus 7 Tablet, considering a constrained scenario. Applying the same classifier, in (Lu and Liu, 2015) the authors achieved 0.03% False Acceptance Rate (FAR) and 0.05% False Rejection Rate (FRR) with a private database. Furthermore, the public HMOG database containing data from the touchscreen and background motion sensors (accelerometer, gyroscope and magnetometer) was presented in (Sitová et al., 2015). The authors achieved 8.50% EER using the OC-SVM classifier. IIn addition, Logistic Regression (Serwadda et al., 2013), Dynamic Time Warping (DTW) (Feng et al., 2014), k-Nearest Neighbours (k-NN) (Antal et al., 2015; Feng et al., 2014) or Random Forest (RF) (Kumar et al., 2016; Mahbub et al., 2016; Saravanan et al., 2014; Shen et al., 2015; Syed et al., 2019) were also broadly used. Another classifier that has been widely used is Binary Support Vector Machine (B-SVM) introduced in (Xu et al., 2014). The difference with the previous classifiers is that it needs to be trained using both genuine and impostor data, unlike the previous classifiers which only genuine data are considered. The authors obtained EER values lower than 1% over a private database acquired using only one device and

under the continuous recognition scenario.

Due to the improvements presented by B-SVM, this classifier has been applied by many studies (Fierrez et al., 2018b; Sharma and Enbody, 2017; Wang et al., 2017; **?**). Using each study their own touchscreen data and experimental protocol, the authors achieved 7.00% EER, 80.0% Area Under the Receiver Operating Characteristic (AUC) and 2.60% EER, respectively. Moreover, studies based on ML demonstrate how adding extra features from the background sensors of the device to the original touchscreen features improves the performance (Acien et al., 2019b; Bo et al., 2014; Sitová et al., 2015). For example, in (Siirtola et al., 2018) the authors achieved on the HMOG database a 7.00% EER when combining touchscreen and accelerometer data.

In recent years, advancements in DL techniques have led to the utilisation of feed-forward ANN as classifiers. Notably, in a study conducted in (Zaliva et al., 2015), a private touchscreen database was used in a constrained scenario, achieving an impressive 99.96% F1-Score using 70% of the data to train. In this work, the authors included two hidden layers, consisting of 50-75 and 30 neurons, respectively. The output layer of the network was equipped with a logistic sigmoid activation function. To preprocess the data and enhance the performance of the classifier, Principal Component Analysis (PCA) was applied, reducing the data's dimensionality. Furthermore, the authors in (Meng et al., 2018b) achieved notable results in an unconstrained scenario using a private touchscreen database. Their approach yielded an impressive Average Error Rate (AER) of 2.40%. The proposed model in their work combined Particle Swarm Optimisation (PSO) with an RBFN (Radial Basis Function Network) classifier, which consisted of three layers: an input layer, a hidden layer, and an output layer. Notably, in the hidden layer, each unit adopted a radial activation function, contributing to the model's effective representation and classification capabilities. These findings highlight the potential of utilising PSO and RBFN-based classifiers in touch-based interaction systems, yielding promising results in

unconstrained scenarios. In addition, LSTM architectures have shown to be well-suited for the task. In (Mao et al., 2022) the authors proposed a 1D-CNN-BiLSTM model that combines the strengths of CNNs and bidirectional LSTMs. The model includes a single convolutional layer with ReLU activation to extract relevant features from the input data. A bidirectional LSTM layer is then employed to capture contextual information in both directions. The model was trained and evaluated using 10-fold cross-validation, ensuring robustness and generalisation. The results highlight the effectiveness of the 1D-CNN-BiLSTM model in touch-based interaction systems. Lastly, in (Acien et al., 2020), a Siamese RNN with two LSTM layers was introduced. The model learns to project embedding vectors to differentiate touch patterns from the same and different subjects. By computing the Euclidean distance between embedding vectors, a performance of 13.00% EER was achieved by training the model with 70% of each subject's swipes. In addition, the authors presented a publicly available database, HuMIdb (Acien et al., 2021a).

As in the case of gait, numerous ML and DL approaches have been proposed in the literature. These approaches have demonstrated the potential for recognising individuals based on their swipe gestures. However, it is evident that further research is required to enhance the performance of swipe biometrics, considering novel configurations and DL architectures. In particular, this thesis explores the potential of novel Transformer architectures for mobile swipe verification.

## 2.4 Sensitive Data in Mobile Biometrics

The automatic processing of subject data acquired by mobile device sensors can reveal a significant amount of personal and sensitive information. In particular, while sensors such as cameras, GPS, or microphones are privacy-sensitive and require explicit subject

permission for use, many other sources such as accelerometers, touchscreens, and/or network connection logs are less protected in terms of privacy. These data can also become crucial in obtaining private subject information (Delgado-Santos et al., 2022a), since they can be processed to ascertain attributes that allow the re-identification of a person, to extract demographic information or data related to their activity and health, among others.

Processing data from which it is possible to extract personal and sensitive information can lead to problems arising from the nature of these data. A common characteristic of sensitive data is in fact its uniqueness for each individual and its strict association to their owner. These implications are particularly relevant with regard to biometric data.

In the biometric scenario, additional risk factors include: the modalities used to store personal data, the owner of the system, the recognition modality (identification or verification), the durability, and class of the traits (physiological or behavioural), etc. These factors can vary the severity of the consequences in terms of privacy and security (Labati et al., 2011). An outline of the different sensitive attributes that can be extracted from mobile sensors is shown in Table 2.5. In the remainder of this Section, examples of the personal and sensitive information extracted from the mobile device sensor data are presented, grouped in several categories depending on the nature of the extracted information and arranged by the particular data acquisition sensor.

### 2.4.1 Demographics

One of the most popular research lines is to predict attributes such as age, gender and ethnicity, which can all be ascribed to the category of demographics. We describe next the different approaches considered in the literature:

**Table 2.5:** Comparison of different state-of-the-art sensitive data acquisition approaches.

| Sensitive Data | Sensors | Study | Classifier | Best Performance |
|---|---|---|---|---|
| Demographics | Motion | (Jain and Kanhangad, 2016) | SVM | Acc. = 76.83% |
| | | (Davarci et al., 2017) | $k$-NN | Acc. = 85.30% |
| | | (Nguyen et al., 2019) | RF | Acc. = 96.00% |
| | | (Singh et al., 2019) | 4 Classifiers | Acc. = 80.00% |
| | | (Sabir et al., 2019) | LSTM + Leave One Out | Acc. = 94.11% |
| | | (Ngo et al., 2014) | HMM | EER = 5.39% |
| | | (Meena and Sarawadekar, 2020) | Ensemble Boosted Tree | Acc. = 96.30% |
| | Touchscreen | (Miguel-Hurtado et al., 2016) | Decision Tree | Acc. = 78.00% |
| | | (Acien et al., 2019a) | AUD | Acc. = 97.00% |
| | | (Nguyen et al., 2019) | RF | Acc. = 99.00% |
| | | (Jain and Kanhangad, 2019) | $k$-NN | Acc. = 93.65% |
| | Network, Location and Application | (Riederer et al., 2015) | Logistic Regresion | Acc. = 72.00% |
| | | (Neal and Woodard, 2018) | RF + Naïve Bayes | Acc. = 91.80% |
| | | (Wu et al., 2019) | XGBoost | Acc. = 80.00% |
| Activity and Behaviour | Motion | (Sun et al., 2010) | SVM | Acc. = 93.2% |
| | | (Anjum and Ilyas, 2013) | Decision Tree | AUC = 99.00% |
| | | (Thomaz et al., 2015) | DBSCAN | Acc. = 76.10% |
| | | (Arnold et al., 2015) | RF | Acc. = 70.00% |
| | | (Chang et al., 2018) | $k$-NN | Acc. = 71.00% |
| | Network, Location and Application | (Wan and Lin, 2016) | Fuzzy Classification | Acc. = 96.00% |
| | | (Chen et al., 2018) | CNN | Acc. = 97.70% |
| | | (Ma et al., 2021) | 2D CNN + RNN | Acc. = 83.00% |

$k$-NN- $k$-Nearest Neighbours, RF- Random Forest, SVM- Support Vector Machine, LSTM- Long-Short Term Memory, HMM- Hidden Markov Model, AUD- Active User Detection, DBSCAN- Density-based Spatial Clustering of Applications, CNN- Convolutional Neural Network, RNN- Recurrent Neural Network, Acc.- Accuracy, EER-Equal Error Rate, AUC- Area Under the Receiver Operating Characteristic.

- **Motion Sensors**: The research conducted in (Davarci et al., 2017) employed accelerometer data during a task where participants were required to tap on different predetermined locations on a device screen. The authors exploited the $k$-NN algorithm, obtaining an age group accuracy of 85.30%. Similarly, in (Nguyen et al., 2019) the authors developed a method to distinguish an adult from a child exploiting the behavioural differences captured by the motion sensors. Based on the hypothesis that children, with smaller hands, will tend to be more shaky, they achieved an accuracy of 96.00% using the RF method. Alternatively, the gender of the subjects was ascertained from walking patterns, which were captured using smartphone motion sensors, as described in (Jain and Kanhangad, 2016). The authors achieved an accuracy of 76.80% by processing with SVM and bag-

ging algorithms. Continuing with gender recognition, the authors in (Meena and Sarawadekar, 2020) presented an approach based on the gait data extracted from smartphone sensors achieving an accuracy of 96.30% using the bagged tree classifier. In addition, in (Singh et al., 2019) the authors obtained an accuracy of 80.00% through PCA from the data extracted from the accelerometer and gyroscope. In (Ngo et al., 2014) the authors focused on extracting gender and age with Hidden Markov Models (HMMs). The authors organised a competition based on accelerometer and gyroscope data acquired by wearable devices, which lead to a percentage error rate of 24.23% for gender and 5.39% for age. With the development of deep learning techniques, it has been possible to achieve better results, as in the case of (Sabir et al., 2019), who obtained an accuracy of 94.11% by using LSTM-RNN, a class of deep learning models particularly useful to capture temporal dependencies underlying in the data.

- **Touchscreen**: Based on swipe and tap gestures, an analysis to identify whether the subject using the device was a child or an adult was performed in (Acien et al., 2019a). For this purpose, an Active User Detection (AUD) algorithm has been used, achieving 97.00% accuracy. In (Tolosana et al., 2022a), a new database of children's mobile interaction was presented. The authors used touch interaction information to classify children into three groups aged 18 months to 8 years old. The authors used a SVM algorithm achieving an accuracy of 90.45%. In addition, the authors in (Nguyen et al., 2019) also conducted a study using RF on tap gestures to distinguish between an adult and a child, achieving an accuracy of 99.00%. Another demographic that can be extracted is gender. An example of this is (Miguel-Hurtado et al., 2016), where from swipe data, the authors achieved a 78.00% accuracy rate using a decision voting scheme from four classifiers: Decision Tree (DT), Naïve Bayes (NB), SVM and LR. Finally, behavioural data from a smartphone's accelerometer, gyroscope and orientation sensors were used in (Jain

and Kanhangad, 2019). The authors used gestural attributes in which the $k$-NN classifier recognises the gender of the subject, providing a classification accuracy of 93.65%.

- **Network, Location and Application**: Studies have shown a strong correlation between a subject's geolocation and usage patterns and their demographics (Almaatouq et al., 2016; Scherrer et al., 2018; Yuan et al., 2012). For instance, in (Riederer et al., 2015) the authors showed how demographic information can be inferred from geo-tagged photos on social networks. Specifically, they performed an analysis of how a person's ethnicity can be extracted from their location. They distinguished between people belonging to three different ethnicity groups with an accuracy of 72.00% using LR. Also, in (Wu et al., 2019) the authors studied how from the spatio-temporal characteristics and geographical context extracted from GPS, it was possible to obtain information on marital status and state of residence with an accuracy of 80.00% based on an XGBoost algorithm. Furthermore, from the gender-related behaviour patterns found in the app, Bluetooth, and Wi-Fi, it is possible to estimate the gender. An example of this is shown in (Neal and Woodard, 2018), where an accuracy of 91.80% is achieved using RF and multinomial NB. This type of contextual behavioural information is used in various thematic services, such as personalisation of advertisements and home screens.

### 2.4.2 Activity and Behaviour

It has been shown that a broad variety of subjects' behaviour or activities can be inferred from mobile device sensor data (Chen et al., 2021):

- **Motion Sensors**: In (Sun et al., 2010) the authors were able to detect whether the person was stationary, walking, running, cycling, climbing stairs, going downstairs

or driving using only the accelerometer information. Their proposed approach, based on SVM, was able to achieve an accuracy of 93.20%. Using accelerometer and gyroscope data, in (Anjum and Ilyas, 2013) the authors developed an application to track the subjects' activities, while the mobile device was kept in their hand, trouser pocket, breast pocket or handbag. Using a DT classifier, they achieved an average AUC curve of over 99.00%. In (Thomaz et al., 2015) the movements made by a subject while eating were estimated by the accelerometer on a smartwatch. In addition, the authors in (Santani et al., 2018) based on smartphone accelerometer data, classified drinking behaviour of young adults. Density-based Spatial Clustering of Applications (DBSCAN) algorithm was used, achieving an accuracy of 76.10%. Even the amount of alcohol taken by subjects can also be extracted from the accelerometer data. One example of this is shown in (Arnold et al., 2015), where if a subject is sober, tipsy or drunk was detected based on the accelerometer data. Their system achieved an accuracy of 70.00% using a RF algorithm. Motion sensors have also been used to extract information related to sleep such as sleep posture and habits. Finally, in (Chang et al., 2018) accelerometer, gyroscope and orientation data from a smartwatch was used to detect sleep posture (supine, left lateral, right lateral, prone) achieving an accuracy over 95.00% calculating the Euclidean distance of the input values. Also, results of over 88.00% accuracy were achieved for the prediction of the hand position (placed on the abdomen, chest or head) using $k$-NN algorithm.

- **Network, Location and Application**: From GPS data, the authors in (Wan and Lin, 2016) determined whether the subject was standing, walking or using other transportation with a fuzzy classifier monitoring the speed and angle of the person obtaining, a matching rate of 96.00% at a five-second interval. Also Wi-Fi can reveal a significant amount of information about subjects' activity as in (Chen et al., 2018). The Wi-Fi Received Signal Strength Indicator (RSSI) was used on

a smartphone to determine what activity subjects were doing, among lying down, falling, walking, running, sitting down, and standing up. An accuracy of 97.70% was obtained with a CNN. Furthermore, the authors in (Ma et al., 2021) used three neural networks on Channel State Information (CSI) measured by Wi-Fi: a 2D CNN as the recognition algorithm, a 1D CNN as the state machine, and an LSTM-RNN as the reinforcement learning agent for neural architecture search. They were able to discriminate whether a person is sitting, standing, walking with an accuracy of 83.00%.

Previous Sections have shown that it is possible to extract sensitive information from background sensors related to the topics studied in this thesis. For a comprehensive analysis of the personal and sensitive information acquired through these sensors, it is highly recommended to consult the survey conducted in (Delgado-Santos et al., 2022a).

## 2.5 Privacy Protection Metrics for Sensitive Data

Privacy protection methods work by modifying the original data in order to deprive it of subject sensitive information. For instance, the modified data should only reveal allowed attributes (e.g., gender) in order to maintain some data utility, in terms of available information, while other attributes (e.g., ethnicity) are suppressed. The degree of privacy achieved is typically related to the extent of data modification; however, the utility of the resulting dataset can be significantly impacted (Garfinkel, 2015). In order to evaluate the effectiveness of privacy protection approaches, the degree of privacy protection achieved, as well as the residual data utility after data modification, should be quantified. The former task can be achieved through specific privacy metrics, whereas the latter can be expressed in terms of reduction of traditional performance metrics such as accuracy or EER.

**Table 2.6:** Some of the most common privacy metrics grouped by the property measured.

| Property | Metric | Input Data |
|---|---|---|
| Anonymity | $k$-Anonymity (Sweeney, 2002) | PAR |
| | $m$-Invariance (Xiao and Tao, 2007) | PAR |
| | *($\alpha$, $k$)*-Anonymity (Wong et al., 2006) | PAR |
| | $\ell$-Diversity (Machanavajjhala et al., 2007) | PAR |
| | $t$-Closeness (Li and Ti, 2007) | PAR, TO |
| | Stochastic $t$-closeness (Domingo-Ferrer and Soria-Comas, 2015) | PAR, TO |
| | *($c$, $t$)*-Isolation (Chawla et al., 2005) | ADE, PAR, TO |
| | *($k$, $e$)*-Anonymity (Zhang et al., 2007) | PAR |
| Differential Privacy | *($d$-$\chi$)*-Privacy (Chatzikokolakis et al., 2013) | PAR, TO |
| | Joint Differential Privacy (Kearns et al., 2014) | PAR, TO |
| | Geo-indistinguishability (Andrés et al., 2013) | PAR, TO |
| | Computational Differential Privacy (Mironov et al., 2009) | ADE, ADR, PAR, TO |
| | Information Privacy (du Pin Calmon and Fawaz, 2012) | ADE, PAR |
| Entropy | Entropy (Shannon, 1948) | ADE |
| | Cross-Entropy (Merugu and Joydeep Ghosh, 2003) | ADE, TO |
| | Cumulative Entropy (Julien et al., 2007) | ADE |
| | Inherent Privacy (Agrawal and Aggarwal, 2001) | ADE, TO |
| | Mutual Information (Lin et al., 2002) | ADE, TO |
| | Conditional Privacy Loss (Agrawal and Aggarwal, 2001) | ADE, TO |
| Success Probability | Privacy Breach (Evfimievski et al., 2004) | ADE, TO |
| | *($d$-$\gamma$)*-Privacy (Rastogi et al., 2007) | ADE, TO |
| | *($\delta$)*-Presence (Nergiz et al., 2007) | ADE, TO |
| | Hiding Failure (Oliveira and Zaıane, 2002) | ADE, TO |
| Error | Euclidean Distance (Shokri et al., 2011) | ADE, TO |
| Accuracy | Confidence Interval Width (Agrawal and Srikant, 2000) | ADE, PAR |
| | *($t$-$\delta$)*-Privacy Violation (Kantarcioğlu et al., 2004) | ADE, PAR, PK, TO |
| | Size of Uncertainty Region (Cheng et al., 2006) | ADE |
| | Customisable Accuracy (Ardagna et al., 2007) | PAR |
| Time | Maximum Tracking Time (Sampigethaya et al., 2005) | ADE |
| | Mean Time to Confusion (Hoh et al., 2007) | ADE, PAR |

ADE - Adversary's Estimate: generally a posterior probability distribution. ADR - Adversary's Resources: computational power, time, etc. PAR - Parameters: for configuring privacy metrics. PK - Prior Knowledge: generally a prior probability distribution. TO - True Outcome: also known as ground truth, it can be used to evaluate the ADE.

Sensitive data acquired through mobile interaction is very heterogeneous and can be *structured*, as in the case of high-level health data, network, location, and application data, or *unstructured*, i.e. motion, position, environmental, touchscreen, and low-level health data. Consequently, different metrics are required depending on the specific application scenario. In this context, we consider data after having undergone modifications in order to suppress or alter specific sensitive attributes, while retaining utility for anal-

ysis and extraction of non-sensitive information. In our discussion, privacy metrics are classified based on their output, in other words, depending on the characteristics of the data that are measured with a specific metric. There is no specific metric that can be applied to every characteristic, so many studies use their own metrics. Table 2.6 shows the metrics considered in our discussion and input data needed for the specific metric computation, grouped by the property measured. According to this criterion, some of the most relevant privacy metrics in the context of data acquired through mobile interaction can be grouped as follows (Wagner and Eckhoff, 2018):

### 2.5.1 Anonymity-based Metrics

Anonymity-based metrics stem from the idea of $k$-anonymity (Sweeney, 2002). These metrics guarantees, upon public release to the community, the inability to differentiate an individual from at least $k-1$ other individuals whose information has also been disclosed. To this aim, only the information disclosed about each individual is used. This is achieved by grouping subject data into equivalence classes with at least $k$ individuals, indistinguishable with respect to their sensitive attributes. $k$-anonymity is independent of the information extraction technique and it quantifies the degree of privacy exclusively considering the disclosed data. It is useful to express the degree of similarity between datasets, namely the original one and the sanitised one, or it can be applied to samples within a single dataset. However, several studies have reported some limitations of $k$-anonymity, which have led to the development of new metrics based on the original, aiming to overcome some of its issues by imposing additional requirements. For instance, $m$-invariance (Xiao and Tao, 2007) modifies $k$-anonymity to allow for multiple and different releases of the same dataset. *($\alpha$,k)*-anonymity (Wong et al., 2006) imposes a predetermined maximum occurrence frequency for sensitive attributes within a class to protect against attribute disclosure. $\ell$-diversity (Machanavajjhala et al., 2007) was

developed to prevent linkage attacks by specifying the minimum diversity within an equivalence class of sensitive information, namely at least $\ell$ well-represented different sensitive values. For a skewed distribution of sensitive attributes, $t$-closeness (Li and Ti, 2007) and stochastic $t$-closeness (Domingo-Ferrer and Soria-Comas, 2015) were introduced, starting from the idea that the distribution of sensitive values in any equivalence class must be close to their distribution in the entire dataset. Consequently, knowledge of the original distribution is needed to compute this metric. Similarly, starting from the original data distribution *(c,t)*-isolation (Chawla et al., 2005) indicates the number of data samples present in the proximity of a sample predicted from the transformed data. Depending on the semantic distance between sensitive subject records, such as in the case of numerical values, $(k,e)$-anonymity (Zhang et al., 2007) requires the range of sensitive attributes in any equivalence class to be greater than a predetermined safe value. Despite the highlighted shortcomings, $k$-anonymity and the derived metrics are still largely employed today in a broad variety of different privacy contexts, but mainly for low-dimensional structured data (Aggarwal, 2005). It has in fact been shown that $k$-anonymity-based properties do not guarantee a high degree of protection in case of high-dimensional data.

### 2.5.2 Differential Privacy-based Metrics

Differential Privacy is a definition that has become popular thanks to its strong privacy statement according to which the data subject will not be affected, adversely or otherwise, by allowing their data to be used in any study or analysis, no matter what other studies, datasets, or information sources, are available (Dwork and Roth, 2014). Differential Privacy is generally achieved by adding noise to the original data. Therefore, in order to quantify the metric as a property of the data indicating the degree of privacy, it is a requirement to have knowledge of the original data. Differential Privacy was defined

in the context of databases to achieve indistinguishability between query outcomes, but thanks to its generality it has found application in different contexts for low-dimensional data, including biometrics and ML systems. It is in fact based on the requirement that independently of the presence of a particular data subject, the probability of the occurrence of any particular sequence of responses to queries is provided by a parameter, $\epsilon$, which can be chosen after balancing the privacy-accuracy trade-off inherent to the system. For a given computational task and a given value of $\epsilon$, there can be several algorithms based on Differential Privacy, which might have different accuracy performances. As in the case of $k$-anonymity, many metrics were originated from the initial definition of Differential Privacy, including Approximate Differential Privacy, which has less strict privacy guarantees, but is able to retain a higher utility (Dwork et al., 2006). $d$-$\chi$-privacy (Chatzikokolakis et al., 2013) allows different measures for the distance between datasets than the Hamming distance used in the definition of Differential Privacy. Joint Differential Privacy (Kearns et al., 2014) applies to systems where a data subject can be granted access to their own private data but not to others'. In the context of location privacy, geo-indistinguishability (Andrés et al., 2013) is achieved by adding Differential Privacy-compliant noise to a geographical location within a determined distance. In contrast to previously described metrics based on Differential Privacy, Computational Differential Privacy (Mironov et al., 2009) adopts a weaker adversary model, favouring accuracy. In order to adopt Computational Differential Privacy, it is necessary to have knowledge of the posterior data distribution reconstructed from the transformed data. Similarly, information privacy (du Pin Calmon and Fawaz, 2012) is met if the probability distribution of inferring sensitive data does not change due to any query output.

### 2.5.3 Entropy-based Metrics

In the field of information theory, entropy describes the degree of uncertainty associated to the outcome of a random variable (Shannon, 1948). Metrics based on entropy are generally computed from the estimated distribution of real data obtained from the sanitised data, even though additional information may be needed for a particular metric, such as the original data or some of the data transformation parameters. When attempting to estimate sensitive information from protected subject data, high uncertainty generally correlates with high privacy. Nonetheless, a correct guess based on uncertain information can still occur. In (Merugu and Joydeep Ghosh, 2003), the degree of privacy protection is quantified by cross-entropy (also referred to as likelihood) of the estimated and the true data distribution in the case of clustered data derived from the original data. A cumulative formulation of entropy was defined in (Julien et al., 2007) in the context of location privacy to measure how much entropy can be gathered on a route through a series of independent zones. Inherent privacy (Agrawal and Aggarwal, 2001) represents another example of metric derived from the definition of entropy, considering the number of possible different outcomes given a number of binary guesses. Mutual information and conditional privacy loss (Agrawal and Aggarwal, 2001; Lin et al., 2002) are also metrics based on entropy. The former provides a measure of the quantity of information common to two random variables and it can be computed as the difference between entropy and conditional entropy, also known as equivocation, which is useful to compute the amount of information needed to describe a random variable, assuming knowledge of another variable belonging to the same dataset. The latter property is built on similar premises, but it considers the ratio between true data distribution and the amount of information provided by another variable revealed.

### 2.5.4 Success Probability-based Metrics

Metrics in this category do not take into account properties of the data but only the outcome of sensitive information extraction attempts, as low success probabilities indicate high privacy. However, even if this trend is observable considering the entire dataset, single subjects' private data could still be compromised. In (Evfimievski et al., 2004), based on the original and estimated data, a privacy breach is defined as the event of the reconstructed probability of an attribute, given its true probability, being higher than a fixed threshold, whereas in (Rastogi et al., 2007), this idea was extended by *(d,γ)*-privacy, in which additional bounds are introduced for the ratio between the true and reconstructed probabilities. In contrast, $\delta$-presence (Nergiz et al., 2007) evaluates the probability of inferring that an individual is part of some published data, assuming that an external database containing all individuals in the published data is available. Hiding Failure (HF) (Oliveira and Zaıane, 2002) is a data similarity metric used to detect sensitive patterns. This metric is computed as the ratio between the sensitive patterns found in the sanitised data set and those found in the original data set. If HF is equal to zero, it means all the patterns are well hidden.

### 2.5.5 Error-based Metrics

Error-based metrics measure the effectiveness of the sensitive information extraction process, for example, using the distance between the original data and the estimate. A lack of privacy generally takes place in case of small estimate errors. In location privacy, the expected estimation error measures the inference correctness by computing the expected distance between the true location and the estimated location using a distance metric, such as the Euclidean distance (Shokri et al., 2011). Furthermore, with particular regard to high-dimensional, unstructured data such as the ones acquired

by mobile background sensors or images, a simple but common approach to quantify privacy consists in comparing the traditional performance metrics of sensitive attribute extraction methods (i.e. accuracy) before and after the data modification process. A significant performance drop is a valid indicator of the effectiveness of a data modification technique.

### 2.5.6 Accuracy-based Metrics

Accuracy-based metrics quantify the accuracy of the inference mechanism, as inaccurate estimates typically show higher privacy. The confidence interval width indicates the amount of privacy given the estimated interval in which the true outcome lies (Agrawal and Srikant, 2000). It is expressed in percentage terms for a certain confidence level. *(t, $\delta$)*-privacy violation (Kantarcioğlu et al., 2004) provides information whether the release of a classifier for public data is a privacy threat, depending on how many training samples are available to the adversary algorithm. Training samples link public data to sensitive data for some individuals, and privacy is violated when it is possible to infer sensitive information from public data for individuals who are not in the training samples. In location privacy, the size of the uncertainty region denotes the minimal size of the region to which it is possible to narrow down the position of a target subject, while the coverage of sensitive region evaluates how a subject's sensitive regions overlap with the uncertainty region (Cheng et al., 2006). A different approach was proposed in (Ardagna et al., 2007). In this work, data subjects are given the possibility to customise the accuracy of the region they are in when submitting it the date to an internet service. The accuracy of the obfuscated region can therefore be seen as an indicator of privacy.

### 2.5.7 Time-based Metrics

Time-based metrics measure the time that elapses before sensitive information can be extracted. For instance, in location tracking, to evaluate a given privacy protection method, it can be useful to measure for how long it is possible to breach privacy by successfully tracking the subject, by computing the maximum tracking time (Sampigethaya et al., 2005) or the mean time to confusion (Hoh et al., 2007).

## 2.6 Privacy Protection Methods for Sensitive Data

Given the amount of personal and sensitive information that can be extracted from mobile device sensors, it is necessary to apply a series of techniques to protect the data, as specified in the GDPR. The data should be used for its primary purpose, consented by the subject, and it should not be possible to obtain additional information from the re-purposed data. Privacy protection methods aim to decrease the effectiveness of information extraction tools by transforming data with regard to specific sensitive attributes, while preserving the utility of the data for the original application scenario. In the remainder of this Section, methods are grouped according to the type of input data they work on: *i) traditional data modification techniques* work well with structured data, as most of them were developed for the purpose of disclosing sanitised databases and their application fulfils the requirements of some of the properties discussed above, thus guaranteeing a certain degree of privacy; *ii) machine learning-based data modification techniques*, which are more apt in the case of complex *unstructured* data, as the relationship between privacy gains and information loss changes completely for high-dimensional, highly correlated unstructured data like images, audio signals and time sequence signals provided by background sensors in mobile devices (Na et al., 2018;

Wieringa et al., 2021). An overview of the different privacy protection methods can be found in Table 2.7.

### 2.6.1 Traditional Data Modification Methods

Traditional data modification techniques have proven to work well with structured data. According to (Verykios et al., 2004), these methods can be divided into the following groups:

- **Data Perturbation** is accomplished by the alteration of an attribute value by a new value. Among traditional data perturbation approaches, randomisation techniques are based on the use of noise to mask the values of the data (Aggarwal and Yu, 2008). By incorporating sufficiently large noise, individual data can in fact no longer be recovered, whilst the probability distribution of the aggregate data can be recovered and used safely from a privacy protection standpoint. Noise can be added to the original values in a number of ways:
  - additive noise, which works by adding a stochastic value to confidential quantitative attributes (Brand, 2002; Mivule, 2013);
  - multiplicative noise, in which protected numerical attributes are multiplied by a stochastic value (Kim and Winkler, 2003);
  - geometric perturbation, in which a mix of additive and multiplicative perturbations are used through a rotation matrix (K. Chen and L. Liu, 2005);
  - nonlinear transformation, applying a sigmoid distortion for mapping the data to a different space but preserving the statistical properties of the data (Bhaduri et al., 2011; Lyu et al., 2018);
  - data condensation, in which the data is transformed into a new distribution

where the new data include the correlations among the different dimension (Aggarwal and Yu, 2008);

– through a combination of the above techniques (Chamikara et al., 2018).

Differential Privacy has been widely used in several applications. For instance, in (Sadhya and Singh, 2016), Differential Privacy was used in a privacy-preserving framework for a recognition system based on soft biometrics, such as age, gender, height, and weight extracted from fingerprints and facial images. In the context of mobile devices, Differential Privacy has also been applied for providing rigorous protection of worker locations in a company-centralised server crowdsensing application (Yang et al., 2018).

- **Data Blocking** replaces an existing attribute value with a predetermined value to indicate the data suppression (it could be "?", "0" or "x" in the case of one-character values) (Karakasidis et al., 2015; Parmar et al., 2011).

- **Data Aggregation or Merging** combines values from a coarser category (Li and Cao, 2012) or processing using a compression algorithm to reduce the number of embedded bits used to store the sensitive data (Ren et al., 2013).

- **Data Swapping** exchanges values of individual records. This technique obtains new data with no valid information making impossible for the adversary to access the real data (Hasan et al., 2016).

- **Data Sampling** releases data of a sample of the population. This technique is based on the conditional probability distribution of the data (Chaudhuri and Mishra, 2006; Liu et al., 2019).

Such strategies have found a large number of different implementations for structured data and are often adopted by governmental or statistical agencies. Many are available

**Table 2.7:** Comparison of different state-of-the-art Privacy Protection Methods for Sensitive Data.

| **Traditional Methods** | | | | | |
|---|---|---|---|---|---|
| **Method/ Classifier** | **Field** | **Sensitive Data Protected** | **Study** | **Best Performance** | **Database** |
| Data Perturbation | Fingerprint<br>Faces Images | Demographics | (Sadhya and Singh, 2016) | 0.45% probability of success<br>@ FAR = 10.00% | VC2002-DB1 Database<br>AR Face Database |
| | Location Data | Location Tracking | (Yang et al., 2018) | TASR ≈ 80.00% | SimpleGeo<br>Places Database<br>Yelp Database |
| Data Blocking | Weather Parameters | Health Parameters | (Parmar et al., 2011) | HF = 0/3 attribute disclosure | UCI Repository:<br>Weather Database |
| Data Aggregation<br>or Merging | Physiologic Signals | Health Parameters | (Ren et al., 2013) | - | MIT-BIH<br>Polysomnographic<br>Database |
| Data Swapping | Personal Attributes | Health Parameters | (Hasan et al., 2016) | l-Diversity = 0 attribute disclosure | UCI Repository:<br>Synthetic Database<br>Adult Database |
| Data Sampling | Personal Attributes | Health Parameters | (Liu et al., 2019) | l-Diversity ≈ 0.15 error | UCI Repository:<br>Adult Database |
| **ML-based Methods** | | | | | |
| **Method/ Classifier** | **Field** | **Sensitive Data Protected** | **Study** | **Best Performance** | **Database** |
| Data Level Methods | | | | | |
| Differential Privacy-based<br>AE | Activity Signals,<br>Biomarkers,<br>Biometric Measures | Health Parameters | (Phan et al., 2016) | Acc. Privacy ≈ 85.00% | Own Database |
| SGD sanititation | Language Modeling | Text Inferring | (McMahan et al., 2017) | -0.13% in accuracy with<br>($4.6e10^{-9}$)-Differential Privacy | Reddit Database |
| Siamese CNN | Face Images | Identity | (Osia et al., 2020) | EER before ≈ 1.00%<br>EER after ≈ 28.00% | IMDB-Wiki +<br>LFW Databases |
| | Activity Signals | Demographics | | EER before ≈ 22.00%<br>EER after ≈ 36.00% | MotionSense Database |
| | Activity Signals | Demographics | (Garofalo et al., 2019) | F1-score SA before = 72.58%<br>F1-score SA after = 52.99% | OU-ISIR Database |
| GAN | Activity Signals | Demographics | (Boutet et al., 2021) | Acc. SA before = 98.50%<br>Acc. SA after = 61.00%<br>Acc. SA before = 98.50%<br>Acc. SA after = 57.00% | MotionSense Database<br>MobiAct Database |
| SAN | Face Images | Demographics | (Mirjalili et al., 2018) | Error Rate SA before = 19.70%<br>Error Rate SA after = 39.30%<br>Error Rate SA before = 8.00%<br>Error Rate SA after = 39.20%<br>Error Rate SA before = 33.40%<br>Error Rate SA after = 72.50%<br>Error Rate SA before = 16.90%<br>Error Rate SA after = 53.80% | CelebA Database<br>MORPH Database<br>MUCT Database<br>RaFC Database |
| | Face Images | Demographics | (Mirjalili et al., 2020) | EER SA before ≈1.00%<br>EER SA after = 20.00%<br>EER SA after = 20.00%<br>EER SA after = 10.00%<br>EER SA after = 10.00% | CelebA Database<br>UTK-face Database<br>MORPH Database<br>MUCT Database |
| Feature Level Methods | | | | | |
| Decision Tree Ensemble | Face Images | Demographics | (Terhörst et al., 2019) | COCR before = 94.70%<br>COCR after = 64.70% | FERET Database |
| | Images | Demographics | (Melzi et al., 2023) | Acc. before = 82.00%<br>Acc.≈ after = 42.00% | FERET Database |
| AE | Face Images | Demographics | (Bortolato et al., 2020) | EER SA before = 1.80%<br>EER SA after = 41.90%<br>EER SA before = 4.90%<br>EER SA after = 41.40%<br>EER SA before = 14.50%<br>EER SA after = 50.20% | CelebA Database<br>LFW Database<br>Adience Database |
| Sensitivity Detector +<br>Triplet Loss | Face Images | Demographics | (Morales et al., 2020) | Acc. SA before = 95.10%<br>Acc. SA after = 54.60% | DiveFace Database |

AE- Autoencoder, SGD- Stochastic Gradient Descent, CNN- Convolutional Neural Network, GAN- Generative Adversarial Network, SAN- Semi-Adversarial Network, FAR- False Acceptance Rate, TASR- Task Assignment Success Rate, HF- Hiding Failure, Acc- Accuracy, EER- Equal Error Rate, AUC- Area Under ROC Curve, SA- Sensitive Attribute, AD- Attribute Disclosure, IVE- Incremental Variable Eliminator, COCR- Correct Overall Classification Rate, LFW- Labeled Faces in the Wild.

in libraries under open-source license, like ARX[1] or the R-package sdcMicro (Templ et al., 2015; Wieringa et al., 2021). However, a critical aspect of these modification techniques is often scalability, i.e. there is a significant performance drop as the number of the dimensions of the database increase; in addition, the computational overhead will increase exponentially with respect to the number of attributes and number of instances. These limitations of the traditional data modification methods are commonly grouped under the label of "curse of dimensionality" (Köppen, 2000).

### 2.6.2 Machine Learning-based Data Modification Methods

In addition to the goal of information extraction as discussed in Section 2.4, considering its potential in big data processing (Qiu et al., 2016), ML-based approaches have in turn been investigated for the purpose of perturbing the data in the attempt to overcome the limitations of traditional modification techniques. Within these algorithms, a two groups subdivision can be made of those that operate at the *i)* data level, and those that operate at the *ii)* feature level, depending on the input data. In this Section we present a brief summary of the most competitive techniques of the two groups according to (Bortolato et al., 2020).

#### 2.6.2.1 Data Level Methods

Algorithms that operate at the data level have raw data as input. The algorithm internally processes the data and generates a transformed database containing the protected sensitive information. Among privacy protection solutions adopted to protect sensitive data in the context of ML models, Differential Privacy-based mechanisms are popular in the literature. In (Phan et al., 2016), a Differential Privacy-based model implemen-

[1] Available at `https://arx.deidentifier.org`.

tation based on perturbing the objective functions was proposed for deep Autoencoders (AE) for human behaviour prediction in a health social network. Such method can be applied to each layer of the network. Similarly, the idea of sanitising the gradient in Stochastic Gradient Descent (SGD) was introduced in (Abadi et al., 2016) for CNN, and for complex sequence models for next-word prediction in (McMahan et al., 2017). Differential Privacy has also been implemented in dedicated Tensorflow[2] and PyTorch[3] libraries. However, Differential Privacy-based mechanisms come at a cost in software complexity, training efficiency and model quality (Tramèr and Boneh, 2020).

Using a convolutional architecture, another possibility is offered by a Siamese architecture, which has two different input vectors while maintaining equal weights in the two halves of the network to acquire comparable output vectors. In (Osia et al., 2020) the authors used this architecture both in the field of facial images, to protect the identity of the person, and in the field of activity recognition to protect the gender of the subject. The authors in (Garofalo et al., 2019) also used a Siamese CNN. In this case their work focused solely on activity recognition while protecting demographic information.

Generative Adversarial Networks (GAN) are also among the most popular techniques considered for this purpose in the literature. GANs are unsupervised methods that exploit two adversarial subnetworks (the *generator* and the *discriminator*), and are able to learn well, in a competitive manner, the statistical structure of high dimensional signals. A GAN-based approach called DySan was developed in (Boutet et al., 2021) for data sanitisation in the context of a mobile application for physical activity monitoring through accelerometer and gyroscope data. Before sending the data to a server hosted in the cloud, gender inferences are prevented by distorting the data while limiting the loss of accuracy on physical activity monitoring.

[2]Available at `https://github.com/tensorflow/privacy`.
[3]Available at `https://github.com/pytorch/opacus`.

A similar approach for privacy protection is based on Semi-Adversarial Networks (SANs). SANs are different from typical GANs in the fact that, in addition to the generator subnetwork, they include two independent discriminator classifiers rather than one. A Semi-Adversarial configuration was proposed in (Mirjalili et al., 2018) for the purpose of image data perturbation. Based on the feedback of two classifiers, where one acts as an adversary of the other, this model was able to hide gender while maintaining the same accuracy in face recognition. The authors extended their work in (Mirjalili et al., 2020), by including, among other things, the possibility of choosing to obfuscate specific attributes (e.g., age and race), while allowing for other types of attributes to be extracted (e.g., gender).

#### 2.6.2.2 Feature Level Methods

There is a second set of methods that, instead of using raw data as input, are applied on the embedded representation of the data. Therefore, a pre-trained model used as a feature extractor is needed. After this stage, this set of features will be the input of the privacy method. Finally, a transformed database that keeps the sensitive data hidden will be the output. An Incremental Variable Eliminations algorithm (IVE) was proposed in (Terhörst et al., 2019). The authors, in training a set of decision trees, obtain a measure of the importance of the variables that predict the sensitive attributes to be reduced. In addition, a modified version of the IVE algorithm was proposed to effectively secure multiple soft-biometric attributes at the same time in (Melzi et al., 2023).

An AE was also used in (Bortolato et al., 2020). The authors introduced the Privacy-Enhancing Face-Representation learning Network (PFRNet), a neural network-based model that works at the level of face representations (templates) from images, aiming to

achieve distinct encodings for both identity and gender in the feature space. The model showed how training a loss function for gender suppression (where the distributions of male and female subjects were similar) for the identity feature space, was an effective way to preserve privacy.

In (Morales et al., 2020) the authors aimed to leave out sensitive information in the decision-making process in an image-based face recognition system without a significant drop of performance by focusing on the feature space. Developed for the purpose of ensuring fairness and transparency, their systems inherently improve the privacy of the data. It works as a independent, decoupled module on top of a pre-trained model and takes as an input the embeddings generated by the model. By defining and minimising its own triplet-loss function, SensitiveNets generates new representations agnostic of gender and ethnicity information, which however still retain information useful for extraction of other attributes.

### 2.6.3 Other Perspectives

Finally, it is important to highlight that in order to protect subjects' privacy while handling their private data, besides data modification methods, other important perspectives to be considered to comply with secure data management practices in relation to privacy include: template protection and data outsourcing.

#### 2.6.3.1 Template Protection

Template protection is an important field of research in the area of biometrics (Melzi et al., 2022a). Templates are compact representations of subjects' biometric data for the purpose of storage. They are transformed into protected biometric references for

security purposes. Template protection schemes should provide the following properties (Nandakumar and Jain, 2015):

- **Irreversibility**: it should be computationally difficult[4] to compute the original template from a subject's protected biometric reference.
- **Revocability**: it should be computationally difficult to compute the original biometric template from multiple instances of protected biometric references derived from the same biometric trait of an individual. Biometric data is permanently associated with the data subject and it cannot be revoked and reissued if compromised, in comparison to tokens or passwords. However, through revocable and irreversible transformations templates can be cancelable, thus mitigating the risks associated with biometric template theft (Patel et al., 2015).
- **Unlinkability**: it should be computationally difficult to determine whether two or more instances of protected biometric reference were obtained from the same biometric trait of a subject. Unlinkability prevents cross-matching across databases.

#### 2.6.3.2 Data Outsourcing

Mobile applications usually exploit cloud resources for model training and inference. Therefore, subjects' personal data containing sensitive information may be store and process by the cloud. If stored on the cloud, data subject privacy undergoes greater risks than being stored locally in the device (Svantesson and Clarke, 2010). Performing the training and inference tasks locally is among alternative solutions investigated. However, the computational resource constraints are much stricter (Chen et al., 2020; Servia-Rodríguez et al., 2017).

[4] A problem is defined computationally difficult if it cannot be solved using a polynomial-time algorithm.

A different approach could be federated learning, a machine-learning strategy according to which models are trained on databases distributed across multiple devices, thus preventing data leakage (Konecny et al., 2016; Lo et al., 2021). However, recent attacks demonstrate that simply maintaining data locality during training processes does not provide sufficient privacy guarantees as intermediate results, if exposed, could still cause some information leakage (Yang et al., 2019). Possible solutions to this problem are given by Differential Privacy mechanisms and Secure Multiparty Computation (SMC) schemes, or a combination of the two (Truex et al., 2019).

Finally, it should be pointed out that the considered techniques should be complemented by widely deployed encryption protocols that would guarantee data security, such as hash functions, secret-key and public-key cryptography, among others (Chi and Zhu, 2017; Hernández-Álvarez et al., 2021).

Directing our attention to the relevant aspects of this thesis, it can be inferred that previous gait verification approaches presented in the literature can preserve specific sensitive attributes (Garofalo et al., 2019), but they require a large volume of labelled data for training.

## 2.7 Conclusions

This Chapter has provided an overview of the main concepts and studies in the literature related to the topics of the thesis. We have first described the different background sensors available in mobile devices together with their main application scenarios. Second, we have described the state of the art on mobile biometric authentication or recognition, in particular for gait and swipe traits. This analysis concludes that most approaches presented in the literature are based on popular CNN and RNN architectures, despite

the limitations described in Section 2.3.1.1. In addition, it is important to remark the high range of performance results depending on the specific experimental protocol, i.e., training and testing the system with the same subjects or not. These aspects motivate the proposal of novel biometric behavioural verification systems based on Transformers and the evaluation of them under realistic experimental conditions.

Furthermore, apparently harmless data can reveal personal and sensitive information about the subject, which must be protected in compliance with the GDPR. We have provided a review of the different kinds of sensitive data that can be extracted by the mobile device sensor data. Furthermore, we have reviewed the most popular privacy metrics that allow a comparison of different aspects and quantify the effectiveness of the privacy protection methods, identifying the most suitable metric for each specific application. Finally, some of the most popular privacy protection methods were also discussed, aiming to offer useful guidelines for managing the trade-off between protecting sensitive attributes while disclosing the allowed attributes, inherent to the privacy problem.

# Part II

# Mobile Biometric Authentication

# Chapter 3

# M-GaitFormer: Mobile Biometric Gait Authentication

## 3.1 Introduction

The popularity of gait recognition has also increased with the success of DL (Filipi Gonçalves dos Santos et al., 2022; Sepas-Moghaddam and Etemad, 2022). Architectures based on CNNs and RNNs, such as LSTM, have proven to be convenient for the task, improving performance and robustness compared to traditional ML techniques. However, these popular DL architectures still have several disadvantages that must be revisited and improved. The main drawbacks are (Hutchins et al., 2022; Vaswani et al., 2017): *i)* sequential computation, not allowing parallelisation within batches; *ii)* compression and condensation of previous time samples, limiting the past information seen, and *iii)* vanishing gradients during back-propagation; the forget gate in a RNN removes a small portion of the previous state after each sample. In order to overcome these limitations, Transformer architectures have been proposed in recent years in several fields (e.g., ma-

chine translation, computer vision, time-series forecasting, etc.) (Tay et al., 2022). Their main advantages in comparison with traditional deep learning architectures are: *i)* they are feed-forward models, processing all sequences in parallel; *ii)* they apply Self-Attention mechanisms, operating over long sequences; *iii)* they process all the sequences efficiently even in one batch; and *iv)* they attend all the previous data simultaneously without the need to summarise them (Vaswani et al., 2017). Recent studies have successfully proved the advantages of Transformers for time-sequential data, outperforming traditional CNN and RNN architectures Li et al. (2021a); Zhang et al. (2022a); Zhou et al. (2021).

To the best of our knowledge, this is the first thesis that intends to explore and propose novel behavioural biometric systems based on Transformers. In particular, this Chapter focuses on gait biometrics and is structured as follows: Section 3.2 explains the main concepts of Transformers, highlighting the key differences between architectures, and proposing a new one, M-GaitFormer. Section 3.3 describes the databases while Section 3.4 provides a description of the system details and experimental protocol. Section 3.5 describes the results achieved in identification scenarios and comparison of Transformers with the state of the art. For completeness, Section 3.6 analyses the application of M-GaitFormer to verification scenarios. Finally, Section 3.7 draws the final conclusions. This Chapter is based on the following publications: (Delgado-Santos et al., 2023a,c).

## 3.2 Methods

Several Transformer architectures have been recently proposed in the literature (Tay et al., 2022; Wen et al., 2022). The original, the Vanilla Transformer, was introduced in 2017 (Vaswani et al., 2017). It was based solely on Self-Attention mechanisms, dispensing with recurrence and convolutions layers entirely. Impressive results were achieved

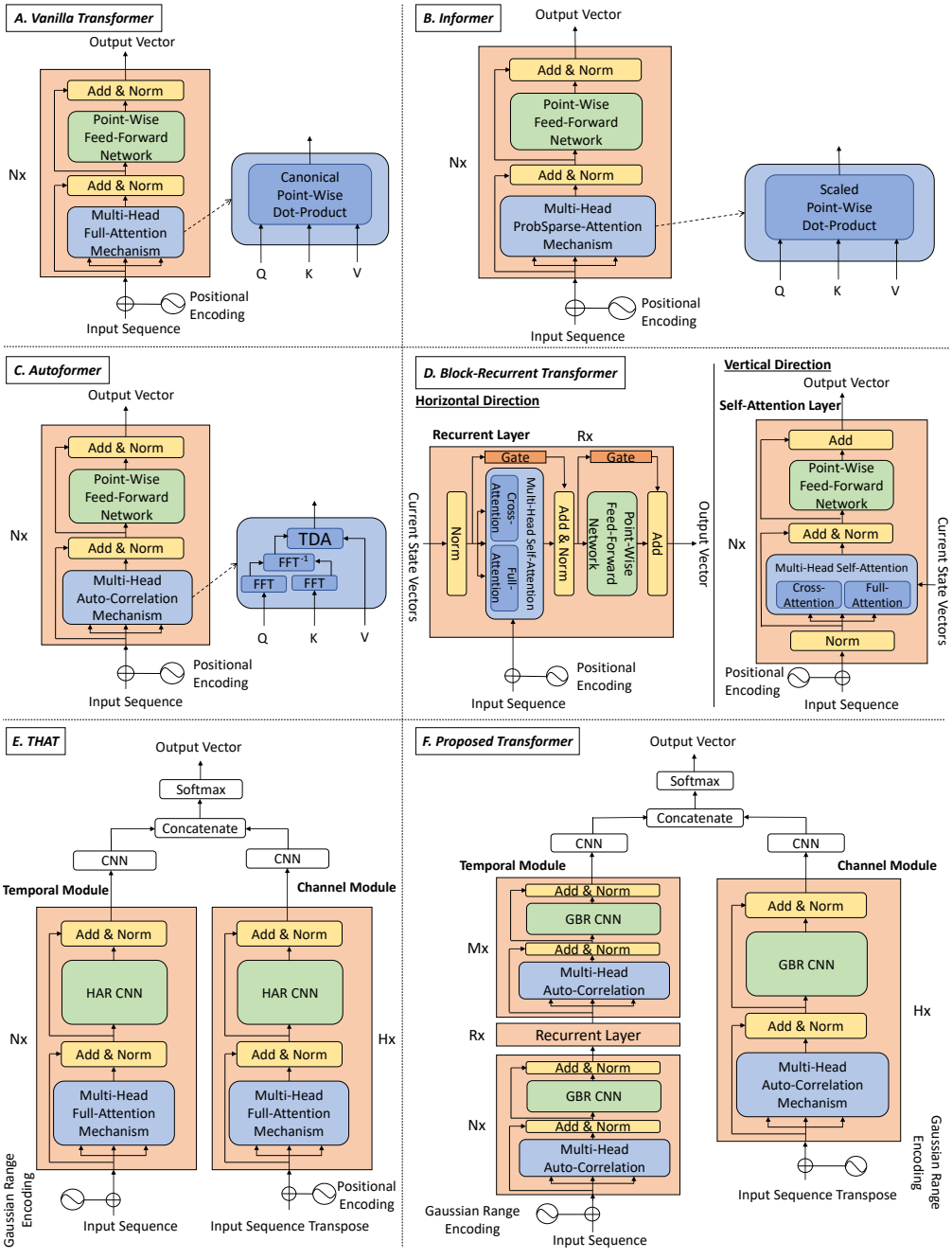
on a machine translation task, reducing the training costs of the best models compared with the literature. Despite these improvements, the Vanilla Transformer has disadvantages for some applications based on time series: *i)* the computational complexity of the attention mechanism is quadratic $O(L^2)$ where $L$ denotes the length of the input sequence; and *ii)* the total memory usage is $O(N \odot L^2)$ where $N$ indicates the number of encoder/decoder layers, limiting the scalability of the model with long sequences. As a result, different Transformer architectures have recently emerged with the aim of addressing the shortcomings of the Vanilla Transformer, including: Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), Block-Recurrent Transformer (Hutchins et al., 2022), and THAT (Li et al., 2021a), amongst others.

This Section provides an overview of the main concepts of Transformers, including the key differences between recent architectures proposed in the literature. To facilitate the understanding of this Section, we include in Fig. 3.1 a graphical representation of the different Transformer architectures. As this Chapter is related to gait recognition, we focus only on the encoder part of the Transformer.

### 3.2.1 Vanilla Transformer

The original Vanilla Transformer was presented in (Vaswani et al., 2017) for the task of machine translation. It was defined as a multi-layer encoder-decoder architecture with no recurrence and convolution layers. Fig. 3.1 *A.* provides a graphical representation of the encoder, which is composed of a stack of $N$ identical layers. Each layer is mainly formed by two different sub-layers: *i)* a multi-head Self-Attention mechanism (Full-Attention), and *ii)* a point-wise feed-forward network. Subsequent of each sub-layer, a residual connection and a layer normalisation are considered ($Add$ & $Norm$ in Fig. 3.1). The input sequence is a matrix $X \epsilon \mathbb{R}^{c \times L}$ where $c$ is the number of channels and $L$ the

**Figure. 3.1:** Graphical representation of the Transformer architectures used in this study (Vanilla Transformer (Vaswani et al., 2017), Informer (Zhou et al., 2021), Autoformer (Wu et al., 2021), Block-Recurrent (Hutchins et al., 2022), THAT (Li et al., 2021a), and our proposed Transformer).




Q: Queries; K: Keys; V: Values; Nx, Hx, Rx, Mx: they refer to the number of layers of each type; FFT: Fast Fourier Transform; TDA: Time Delay Aggregation; HAR CNN: Human Activity Recognition CNN; GBR CNN: Gait Biometric Recognition CNN.

length of the sequence.

The encoder maps each sample $l$ of the input sequence $X = (x_0, x_1, ..., x_l, ..., x_L)$ into hidden states $Z = (z_0, z_1, ..., z_l, ..., z_L)$. The output of each sub-layer is $LayerNorm(X + sublayer(X))$, where $sublayer(X)$ is the function implemented by the multi-head Self-Attention mechanism (Full-Attention) or the point-wise feed-forward network. Both the input $X$ and output $Z$ have the same dimension $L$ to facilitate the work of the residual connections. As no recurrence and convolutional layers are considered in the Vanilla Transformer, a previous encoding of the model is needed to keep certain information about the position of the sample $l$ in the input sequence. This is achieved using a positional encoding placed at the input of the model.

We describe next the key aspects of the positional encoding, multi-head Self-Attention mechanism (Full-Attention), and the point-wise feed-forward network for a better understanding of the Vanilla Transformer, and the later Transformer implementations.

#### 3.2.1.1 Positional Encoding

This stage encodes the relative and/or absolute position $pos$ of the sample $l$ in the input sequence. In the original work (Vaswani et al., 2017), the authors preserved the relative context using a fixed point encoding with the sine and cosine functions:

$$PE_{(pos,2l)} = sin(pos/10000^{2l/L}) \tag{3.1}$$

$$PE_{(pos,2l+1)} = cos(pos/10000^{2l/L})$$

where $L$ is the total length of the input sequence. The positional encoding has the same length $L$ as the embeddings, so that the two can be summed. The output of the

positional encoding is:

$$\hat{x_l} = x_l + PE_{(l)} \tag{3.2}$$

#### 3.2.1.2 Multi-Head Self-Attention Mechanism

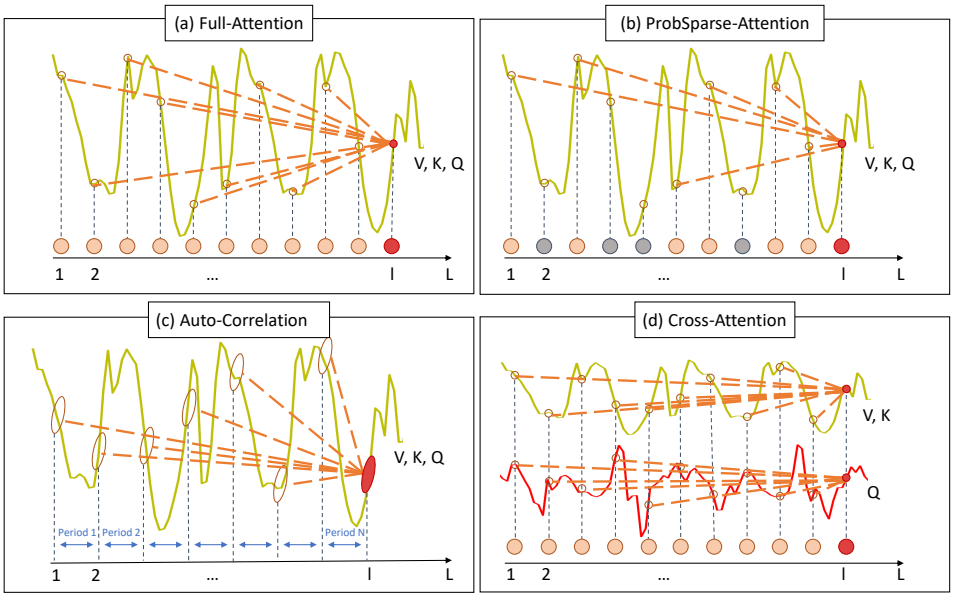
This mechanism is responsible for mapping scattered points along the entire sequence, studying the long-range dependencies. This mechanism avoids the limited time window problem of previous architectures (e.g., RNNs). The information aggregation is accomplished with a Full-Attention mechanism where the outputs are the weighted sum of the values $V$ according to the canonical point-wise dot-product of the queries $Q$ with the corresponding keys $K$. Fig. 3.2 (a) provides a graphical representation of the Full-Attention mechanism. The solid line represents the input sequence with its values $V$, keys $K$, and queries $Q$. The red point represents the sample $l$ in the sequence with length $L$. The orange points are the scattered points mapped in the Full-Attention mechanism for the red point at sample $l$. The Full-Attention mechanism can be defined as:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3.3}$$

where $d_k$ is the dimension of the queries $Q$ and keys $K$, and $\sqrt{d_k}$ is a scaling factor that enables flatter gradients. $Q = XW_Q$, $K = XW_K$, and $V = XW_V$ are the linear projections of $X$ in the corresponding projection parameters $d_k$, $d_k$, and $d_v$ respectively where $W_Q \, \epsilon \, \mathbb{R}^{L \times d_k}$, $W_K \, \epsilon \, \mathbb{R}^{L \times d_k}$, and $W_V \, \epsilon \, \mathbb{R}^{L \times d_v}$. The computational cost is quadratic $O(L^2)$ where $L$ denotes the length of the input sequence.

Alternatively to apply one single projection of the queries, keys, and values, better results can be achieved with $h$ independent projections to $d_k$, $d_k$, and $d_v$ respectively. The multi-head Self-Attention is based on a concatenation and final projection of the $h$

**Figure. 3.2:** Graphical representation of Attention and Auto-Correlation mechanisms. (a) Full-Attention (Vanilla Transformer (Vaswani et al., 2017)); (b) ProbSparse-Attention (Informer (Zhou et al., 2021)); (c) Auto-Correlation (Autoformer (Wu et al., 2021)); and (d) Cross-Attention (Block-Recurrent Transformer (Hutchins et al., 2022)).




The solid line represents the input sequence and the red one (second line) the current states in Cross-Attention. The red points/series are the sample $l$ of the sequence of length $L$ with $V$ values, $K$ keys, and $Q$ queries. The orange points represent the mapped points/series along the entire sequence, while the grey ones are points not mapped. Figure adapted from (Wu et al., 2021).

independent heads:

$$MultiHead(Q, K, V) = [head_1, ..., head_h]W^O \tag{3.4}$$

where $head_i = Attention(Q_i, K_i, V_i)$ and $W^O \; \epsilon \; \mathbb{R}^{hd_v \times L}$ is the final attention matrix. To achieve the same length $L$ of the input sequence, $d_v = L/h$. Therefore the attention matrix of Full-Attention is $L \times L$.

#### 3.2.1.3 Point-Wise Feed-Forward Network

In addition to the multi-head Self-Attention sub-layer, the Vanilla Transformer has a point-wise feed-forward network. This consists of two linear transformations with a Rectified Linear Unit (ReLU) activation in between, operating in each position independently. The input and output dimensions are the same, $L$.

To summarise, the Vanilla Transformer has shown great advances in Natural Language Processing and Computer Vision applications but, as commented before, still needs to be adapted for time sequences. Aspects such as the periodicity or seasonality, and long- and short-range dependencies still need to be revisited (Wen et al., 2022). To alleviate these drawbacks, different Transformers have been proposed in the research community, modifying aspects such as the multi-head Self-Attention sub-layer and the positional encoding. We describe next the most popular methods.

### 3.2.2 Informer

In (Zhou et al., 2021) a new Transformer architecture named Informer was presented. Informer is an adaptation of the Vanilla Transformer for Long Sequence Time-series Forecasting (LSTF). Some limitations of the Vanilla Transformer are the quadratic time complexity $O(L^2)$ and the high memory usage $O(L^2)$ for each encoder layer; and the inherent limitation of the encoder-decoder architecture. To overcome these drawbacks, the authors proposed several improvements. The multi-head Self-Attention mechanism based on Full-Attention was changed by ProbSparse-Attention to scattered points, as provides Fig. 3.1 *B*. The Full-Attention to the input sequence is reduced to half, more favourable handling long-range sequences. The canonical dot-product was replaced by a scaled dot-product. Informer reduces the time complexity to $O(L\ log\ L)$ and the memory

usage to $O(L\ log\ L)$ for each layer. In addition, previous studies have shown a potential sparsity in Full-Attention. As a result, the authors decided to use a selective strategy on all probabilities, i.e., Sparse-Attention (Child et al., 2019) (sparsity coming from separate spatial correlations) and LogSparse-Attention (Li et al., 2019) (selecting points through exponentially increasing intervals). Fig. 3.2 (b) provides a graphical representation of the ProbSparse-Attention mechanism. The solid line denotes the input sequence with the extracted values $V$, keys $K$, and queries $Q$. The red point represents the sample $l$ in the input sequence. The ProbSparse-Attention mechanism, unlike the Full-Attention mechanism that looks at all previous points, chooses selected dominant points (orange) in the input sequence, while the grey ones are not used.

### 3.2.3 Autoformer

Autoformer was presented in (Wu et al., 2021) for the task of long-term forecasting. In this Transformer architecture, the original multi-head Self-Attention mechanism based on Full-Attention was changed by Auto-Correlation. Contrary to previous Transformers, where the proposed dot-product only establishes point connections, the Auto-Correlation mechanism not only utilises long-range dependencies but also periodicity-based dependencies. Using series-wise instead of point-wise connections, Autoformer achieves $O(L\ log\ L)$ time complexity and $O(L\ log\ L)$ memory usage for each layer, and breaks the information utilisation bottleneck. Fig. 3.2 (c) shows a graphical representation of Auto-Correlation. It takes into consideration series of points in the same position during previous periods of the input sequence instead of scattered points.

Fig. 3.1 *C.* provides a graphical representation of Autoformer. The multi-head Auto-Correlation sub-layer comprises two main sub-blocks: *i)* an aggregated top-k similar sub-series, calculated by Fast Fourier Transform (FFT) and based on periodicity (instead of

scattered points like the Self-Attention family), and *ii)* Time Delay Aggregation (TDA) among periods (instead of point-wise dot-product like in the Self-Attention family), used for the information aggregation.

The *aggregated top-k similar sub-series* presents series-wise connections based on period-based dependencies. The sub-series are correlated between them at the same position in previous periods, which are congenitally sparse. For an input sequence $X = (x_0, x_1, ..., x_l, ..., x_L)$, $X \ \epsilon \ \mathbb{R}^{c \times L}$ where $c$ is the number of channels and $L$ the length of the input sequence, the Auto-Correlation $R_{XX}(\tau)$ can be obtained by FFT based on Wiener-Khinchin theorem as:

$$\begin{aligned} S_{XX}(f) &= FFT(X)FFT^{*}(X) \\ R_{XX}(\tau) &= FFT^{-}1(S_{XX}(f)) \end{aligned} \tag{3.5}$$

where $FFT^{*}$ is the conjugate operation, $FFT^{-1}$ its inverse, and $S_{XX}(f)$ is the Auto-Correlation obtained in the frequency domain.

The *TDA* sub-block links the sub-series over the selected time delays $\tau_1, ..., \tau_k$. This operation aligns sub-series in the same phase of the predicted periods, contrary to point-wise dot-product in the Self-Attention family. Finally, the sub-series are aggregated by softmax normalised function. The Auto-Correlation mechanism can be defined as:

$$\tau_1, ..., \tau_k = \underset{\tau \ \epsilon \ (1,...,L)}{argTopK}(R_{Q,K}(\tau))$$

$$\hat{R}_{Q,K}(\tau_1), ..., \hat{R}_{Q,K}(\tau_k) = SoftMax(R_{Q,K}(\tau_1), ..., R_{Q,K}(\tau_k)) \tag{3.6}$$

$$Auto - Correlation(Q, K, V) = \sum_{i=1}^{k} Roll(V, \tau_i)\hat{R}_{Q,K}(\tau_i)$$

where $argTopK$ takes the output of $topK$ Auto-Correlations along $l$, $R_{Q,K}$ is the Auto-Correlation between $Q$ and $K$ series, and $Roll(V, \tau_i)$ scrolls $X$ with a $\tau$ time delay, re-introducing the elements moved beyond the first position to the last one.

### 3.2.4 Block-Recurrent Transformer

The Block-Recurrent Transformer was introduced in (Hutchins et al., 2022) for the task of auto-regressive language modelling. This Transformer introduces a recurrent form of attention. It is presented as an alternative to using the dot-product or periodicity-based series mechanism, which fix an attention window size. The Block-Recurrent Transformer summarises the sequence that the model has previously seen. The time complexity is linear $O(L)$ for each layer. The recurrent layers operate on series-wise connections as in the Autoformer, achieving linear memory consumption $O(L)$ in each layer. The Block-Recurrent Transformer is based on a sliding-window attention mechanism (Beltagy et al., 2020). Given an input $X$ with length $L$, a causal mask is applied by a sliding window with size $W$ where every sample can attend only to the previous $W$ samples. Being the attention matrix of Full-Attention $L \times L$, the Block-Recurrent Attention matrix is $W \times W$, where $W << L$. The sliding-window attention processes multiple blocks of size $W$ at the same time.

Fig. 3.1 *D.* provides a graphical representation of the Block-Recurrent Transformer architecture, which comprises two main directions: *i)* vertical direction (Self-Attention Layer in Fig.3.1 *D.*), where layers are placed in the usual way; and *ii)* horizontal direction (Recurrent Layer in Fig.3.1 *D.*), where layers contain recurrence. Both directions attend to the input sequence $X$ and to the current states $S$.

The *vertical direction* presents a multi-head Self-Attention sub-layer with two attentions: *i)* Full-Attention to the input sequence $X$ as shown in Fig. 3.2(a); and *ii)*

Cross-Attention applied in a similar way to the original Vanilla Transformer (Vaswani et al., 2017), with the main difference being that the queries $Q$ come from the current states $S$, which are initialised to 0, whereas the keys $K$ and values $V$ are extracted from the input sequence $X$, Fig. 3.2(d).

The *horizontal direction* also presents a multi-head Self-Attention sub-layer with two attentions: *i)* Cross-Attention to the input sequence $X$ to extract the queries $Q$ while the keys $K$ and values $V$ are extracted from the current states $S$, Fig. 3.2(d), and *ii)* Full-Attention to the current states $S$, Fig. 3.2(a). The horizontal direction applies recurrence where the residual connections are replaced by gates, allowing the model to forget. Also, the gates help the model to apply Full-Attention and Cross-Attention in parallel. For the recurrence, the current states $S$ are modified by residual connection gates. The input of the state at the next window ($s_{w+1}$) depends on the output of the state at the actual window ($s_w$):

$$
\begin{aligned}
s_{w+1} &= s_w \odot g + z_w \odot (1 - g) \\
g &= \sigma(b^{(g)}) \\
z_w &= W^{(z)} h_w + b^{(z)}
\end{aligned}
\tag{3.7}
$$

where $\odot$ is the point-wise multiplication, $g$ the gate, $z_w$ the learned convex combination, $b^{(g)}$ and $b^{(z)}$ are trainable bias vectors (learned functions between the distance of the query $Q$ and key $K$), $W$ the weight matrix, $h_w$ the output of the corresponding sub-layer (i.e., multi-head Self-Attention mechanism or point-wise feed-forward network), and $\sigma$ the sigmoid function.

The Block-Recurrent Transformer applies layer normalisation before the multi-head

Self-Attention sub-layer, and before the point-wise feed-forward network. Dropout is also introduced before the multi-head Self-Attention sub-layer and after the point-wise feed-forward network.

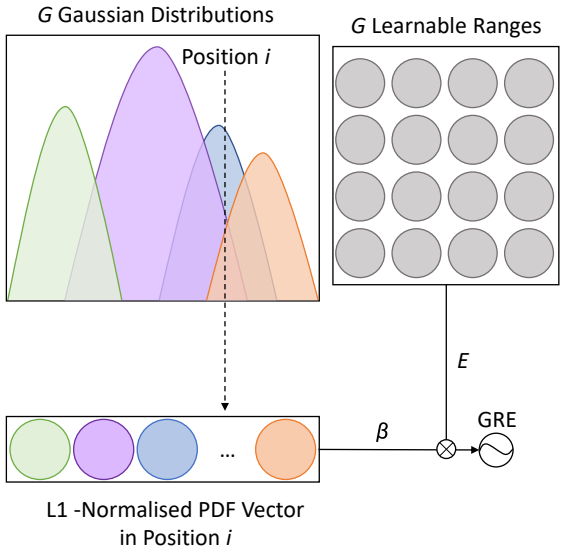
### 3.2.5 THAT

Contrary to images, which have spatial information in two-Dimensions (2D), temporal sequences might consider spatial information in one-Dimension (1D) in each time position. Furthermore, they can extract temporal information for each time position in a second dimension. The spatial information is available in the same way, between the different channels of each time sample, which can be called as channel-over-time features. On the contrary, being a temporal sequence, there are time-over-channel features, which need to be treated as a temporal sequence.

Based on this idea, the *Two-stream Convolution Augmented Human Activity Transformer* (THAT) model was proposed in (Li et al., 2021a). The authors proposed a new Transformer architecture for Human Activity Recognition (HAR). Fig. 3.1 *E.* provides a graphical representation of the THAT Transformer. The model contains two parallel modules for the feature extraction: *i)* Temporal Module (in charge of time-over-channel features), and *ii)* Channel Module (in charge of channel-over-time features). Subsequently, all extracted features are concatenated for the prediction task.

The authors claimed that the original positional encoding considered in the Vanilla Transformer (Vaswani et al., 2017) might not be sufficient to capture all the temporal information along the sample as it is defined on a single point. As a result, the authors proposed a Gaussian Range Encoding (GRE), suggesting the use of a range of points rather than just one. As shown in Fig. 3.3, several ranges $G$ can be used at the same time, allowing to have different contexts of the sample $x_l$.

**Figure. 3.3:** Graphical representation of the Gaussian Range Encoding (GRE). PDF: Probability Density Function.




PDF: Probability Density Function.

Assuming $G \epsilon \mathbb{R}^G$ different ranges, $\mathcal{N}(\mu^G, \sigma^G) \epsilon \mathbb{R}^{L \times G}$ is a Gaussian distribution with the probability $p^G(l)$. Being $p_l = (\frac{p^1(l)}{\zeta}, ..., \frac{p^G(l)}{\zeta})$ the distribution over the $G$ ranges with a normalisation factor $\zeta$, $V = (v_1, ..., v_G)$ is the values vector over the ranges. All $\mu$, $\sigma$, and $V$ variables are initialised randomly and re-adjusted with the training of the whole model. To summarise, the output of the GRE at the position of sample $l$ is:

$$\hat{x_l} = x_l + V^T p_l \tag{3.8}$$

In addition, as the point-wise feed-forward layer proposed in the Vanilla Transformer (Vaswani et al., 2017) focuses attention on a single point in time, the authors implemented a multi-scale CNN with adaptive Scale-Attention in both Temporal and Channel Modules. They replaced the linear transformations of the original feed-forward layer with a HAR CNN. Also, by introducing Scale-Attention Adaptive, the training can

be adjusted to the different ranges introduced by the GRE.

Finally, THAT has quadratic time complexity $O(L^2)$ and the high memory usage $O(L^2)$ for each encoder layer, since the model uses Self-Attention (i.e., Full-Attention similar to the Vanilla Transformer).

### 3.2.6 Proposed Transformer: M-GaitFormer

Finally, Fig. 3.1 *F.* presents our new proposed Transformer based on a selection of the best components presented in previous Transformer architectures. First, we consider a parallel two-stream architecture with Temporal and Channel Modules, similar to the THAT approach presented in (Li et al., 2021a). Unlike the THAT model, we consider a GRE as input of both Temporal and Channel Modules. In addition, for the Temporal Module (left branch), we consider a combination of multi-head Auto-Correlation layers, proposed in Autoformer (Wu et al., 2021), and a recurrent layer in between, proposed in Block-Recurrent Transformer (Hutchins et al., 2022). For the multi-head Auto-Correlation layer, we design a specific multi-scale Gait Biometric Recognition (GBR) CNN sub-layer. Regarding the Channel Module (right branch), we consider a multi-head Auto-Correlation sub-layer together with a multi-scale GBR CNN sub-layer. After each sub-layer, a residual connection is applied followed by a normalisation of the layer, similar to the Vanilla Transformer (Vaswani et al., 2017). The time complexity and memory usage of each layer with Auto-Correlation is $O(L\ log\ L)$, whereas for the recurrent layer this is $O(L)$.

## 3.3 Databases Description

Two popular public databases used for research in gait recognition on mobile devices are considered in the evaluation framework of the present study: *i)* whuGAIT (Zou et al., 2020), and *ii)* OU-ISIR (Ngo et al., 2014). These databases have been selected as they also contain predefined experimental protocols for the identification task (i.e., development and evaluation datasets), allowing for a fair comparison between existing state-of-the-art approaches.

### 3.3.1 WhuGAIT Database

The whuGAIT database was introduced in (Zou et al., 2020). This database comprises accelerometer and gyroscope data acquired using Samsung, Xiaomi, and Huawei smartphones in unconstrained scenarios. The sampling frequency of the accelerometer and gyroscope sensors is 50 Hz. A total of 118 subjects participated in the acquisition, and both walking and non-walking sessions were considered.

The database is divided in 8 subsets, being the subset #3 the one used in this study. It comprises data of all 118 users in both development and evaluation datasets. Following the experimental protocol presented by Tran *et al.* in (Tran et al., 2021), time sequences of 80 gait signals are used. Being the sampling rate 50 hz, some important information can be missing. For this reason, an overlapping of 97% in the development set is implemented, while the testing set remains without overlapping. From the development dataset, 1500 random samples are used for training, while the remaining samples are included in the validation set. The entire evaluation dataset is considered for testing.

### 3.3.2 OU-ISIR Database

The OU-ISIR database was presented in (Ngo et al., 2014). This database comprises 745 subjects; the largest public mobile device gait biometric database to date. Data from accelerometer and gyroscope sensors were collected using three IMUs and a Motorola ME860 smartphone around the waist of the subject. The sampling frequency of the sensors is 100 Hz. Subjects had to perform 4 different activities (two flat walking, slope-up walking, and slope-down walking). The database is divided into two different subsets. The first subset includes data from 744 users collected by one IMU located in the middle of the subject's back at waist-height. The second one contains data from 408 subjects collected by the three IMUs and the smartphone.

## 3.4 Experimental Setup

### 3.4.1 System Details

This Section provides the system configuration details of Transformers and traditional DL architectures (i.e., CNNs and RNNs) considered in the experimental framework of this thesis.

The same inputs to the models are used for all approaches. For the whuGAIT database, a total of 80 time signals (around 1.5 seconds each) are extracted from the 3-axis accelerometer and gyroscope sensors following the approach presented in (Tran et al., 2021). Also, following the same experimental protocol, we consider an overlapping of 97% between samples in training. For the OU-ISIR database, 128 time signals (around 1.5 seconds each) are extracted from the 3-axis accelerometer and gyroscope sensors following the approach presented in (Zou et al., 2020). Also, following the authors

suggestions, we consider an overlapping of 61% between samples in training.

For a better comparison of Transformer architectures with popular DL architectures, we consider the following approaches: *i)* CNNs, *ii)* RNNs, and *iii)* a hybrid configuration based on the combination of CNNs and RNNs. These DL models are widely considered for gait biometric recognition, achieving state-of-the-art results. CNNs have shown advantages in capturing spatial dependencies, while RNNs are better to capture the temporal dependencies.

We provide next a description of the optimal parameters of the networks:

- *CNN*: we consider four 1D convolutional layers with 6 units each and kernel size 5, followed by one dense layer with $\frac{3}{2}L$ units (where $L$ is the length of the time sequence), and one softmax layer. After every 2 convolutional layers, we use max-pooling and dropout with a 0.5 rate. ReLU activation functions are used in both convolutional and dense layers.

- *RNN*: we consider three LSTM layers with 3 units each followed by one dense layer with $\frac{3}{2}L$ units, and one softmax layer.

- *CNN-RNN*: it comprises two parallel modules, *i)* four convolutional layers with 6 units each and kernel size 5, and *ii)* three LSTM layers with 3 units each. After both modules, a feature concatenation is applied, followed by one dense layer with $\frac{3}{2}L$ units, and one softmax layer. We also consider dropout with 0.5 rate after each convolutional layer.

- *Vanilla Transformer (Vaswani et al., 2017)*: we consider the positional encoding together with the encoder part of the Vanilla Transformer. The model consists of $N = 5$ layers. For the multi-head Self-Attention sub-layer, 8 heads are considered with Full-Attention whereas for the point-wise feed-forward network we consider

two linear layers (layer 1 with $L$ units and layer 2 with $L * 4$ units) with ReLU activation and dropout in between.

- *Informer (Zhou et al., 2021)*: we consider the same structure as the Vanilla Transformer but changing in the multi-head Self-Attention sub-layer the Full-Attention to ProbSparse-Attention. The model is composed of $N = 5$ layers. For the multi-head Self-Attention sub-layer, 8 heads are considered whereas for the point-wise feed-forward network we consider two linear layers (layer 1 with $L$ units and layer 2 with $L * 4$ units) with ReLU activation and dropout in between.

- *Autoformer (Wu et al., 2021)*: the same structure as the Vanilla Transformer is considered but changing the Self-Attention mechanism for the Auto-Correlation mechanism. The model comprises $N = 5$ layers with 8 heads in the multi-head Auto-Correlation sub-layer. For the point-wise feed-forward network we consider two linear layers (layer 1 with $L$ units and layer 2 with $L * 4$ units) with ReLU activation and dropout in between.

- *Block-Recurrent Transformer (Hutchins et al., 2022)*: it comprises 12 layers: $N = 9$ multi-head Self-Attention layers with Cross-Attention and Full-Attention (8 heads), followed by $R = 1$ recurrent layer, and $M = 2$ more multi-head Self-Attention layers with Cross-Attention and Full-Attention (8 heads). In each layer, the point-wise feed-forward network is composed of two linear layers (layer 1 with $L$ units and layer 2 with $L*4$ units) with ReLU activation and dropout in between.

- *THAT (Li et al., 2021a)*: this is a two-stream convolution Transformer architecture. In the first stream (Temporal Module) the time-over-channel features are analysed. To this aim, a GRE is used together with the original multi-head Self-Attention sub-layer (Full-Attention with 8 heads). The HAR CNN sub-layer is based on a multi-scale CNN (3 convolutional layers with $L$ units each, ReLU activation functions, and kernel sizes 1, 3, and 5 respectively, followed by dropout

layers). The Temporal Module contains $N = 9$ layers. For the second stream (Channel Module) the data is transposed to extract the channel-over-time features, adopting the original Vanilla Transformer structure with positional encoding. The multi-head Self-Attention sub-layer contains Full-Attention with 6 heads. The HAR CNN sub-layer is based on a multi-scale CNN (3 convolutional layers with $L$ units each, ReLU activation functions, and kernel sizes 1, 3, and 5 respectively, followed by dropout layers). The Channel Module contains $H = 1$ layer.

- *Proposed Transformer*: we consider a two-stream Transformer based on Temporal and Channel Modules. Both modules use a GRE. The Temporal Module comprises 12 layers: $N = 9$ multi-head Auto-Correlation layers (8 heads), followed by $R = 1$ recurrent layer (8 heads), and $M = 2$ multi-head Auto-Correlation layers (8 heads). In each layer, the GBR CNN sub-layer is based on a multi-scale CNN (4 convolutional layers with $L$ units each, ReLU activation functions, and kernel sizes 1, 3, 5, and 7 respectively, followed by dropout layers). The Channel Module comprises $H = 1$ layers. In all of them we consider multi-head Auto-Correlation mechanism with 6 heads. The GBR CNN sub-layer is based on a multi-scale CNN (4 convolutional layers with $L$ units each, ReLU activation functions, and kernel sizes 1, 3, 5, and 7 respectively, followed by dropout layers). These parameters have been selected according to the performance achieved with the proposed Transformer as documented in Section 3.5

For the training of the models, we use cross-entropy and Adam optimiser with default parameters (learning rate of 0.001). All models are adapted to the gait biometric recognition task. To this aim, after the models we include 2 convolutional layers ($L$ units each, ReLU activation functions, and kernel sizes 128, followed by dropout layers) with max-pooling and a linear layer with softmax activation function. For the THAT and proposed Transformer, we also consider feature concatenation of the Temporal and

Channel Modules as described in Fig 3.1 *E.* and *F.*

### 3.4.2 Experimental Protocol

#### 3.4.2.1 WhuGAIT Database

In the experimental protocol of the whuGAIT database, as proposed in (Zou et al., 2020), a predefined division of the database into development and evaluation datasets was implemented to facilitate comparison among different approaches. Each subject's data was partitioned, allocating 90% of the samples for development and the remaining 10% for final evaluation. This resulted in 33,104 samples considered for the development dataset and 3,740 samples for the final evaluation.

The gait curve was divided using a fixed time length, specifically, samples were collected with a time interval of 2.56 seconds. With a data collection frequency of 50Hz, each sample had a length of 128. To enrich the dataset, an overlap of 1.28 seconds was introduced. In total, 29,274 samples were collected, with 26,283 samples designated for training and the remaining 2,991 for testing. This meticulous approach enhances the robustness of the dataset and ensures a comprehensive evaluation of the proposed methodology.

#### 3.4.2.2 OU-ISIR Database

In the experimental protocol of the OU-ISIR database, we adopted the predefined division into development and evaluation datasets proposed in (Zou et al., 2020). For each subject, 87.5% of the samples were allocated for development, with the remaining 12.5% reserved for final evaluation. This partitioning strategy involved 13,212 samples for the development dataset, while the remaining 1,409 samples were designated for the final

evaluation, all collected from 745 subjects.

Despite the database's richness in subjects, there is a scarcity of inertial gait data for individual subjects, ranging from 5.61 seconds to a maximum of 18.73 seconds, with a sensor frequency of 100Hz. The inclusion of this database aimed to assess the performance of deep neural networks in scenarios with limited training samples and a large number of subjects. Data was sampled from the inertial sequence with a length of 128 time points and an interval of 50, resulting in a data sample dimension of $6 \times 128$.

To address the limited training samples, a partitioning strategy of 7:1 for training and testing was employed, ensuring a minimum of 8 data samples for any single subject. This yielded 13,212 samples for training and 1,409 samples for testing.

## 3.5 Experimental Results

This Section analyses the performance of the different state-of-the-art Transformer architectures considered in this study (i.e., Vanilla, Informer, Autoformer, Block-Recurrent Transformer, THAT, and our proposed M-GaitFormer architecture) for the topic of gait biometric identification on mobile devices. Section 3.5.1 provides a comparison of Transformer architectures with traditional DL architectures such as CNNs and RNNs. Finally, Section 3.5.2 provides a comparison of the proposed M-GaitFormer with the state of the art.

### 3.5.1 Experiment 1: Transformers vs. Traditional DL Architectures

Table 3.1 provides a comparison of traditional DL models and recent Transformers for the whuGAIT and OU-ISIR databases. The best results achieved for each database

**Table 3.1:** Comparison in terms of Rank-1 accuracy of traditional DL models (CNN, RNN) and recent Transformers for biometric gait identification.

| **Model** | | **Database** | |
|---|---|---|---|
| | | **whuGAIT** | **OU-ISIR** |
| CNN | | 75.31% | 32.51% |
| RNN | | 82.42% | 44.15% |
| CNN + RNN | | 84.54% | 46.63% |
| Vanilla Transformer (Vaswani et al., 2017) (Positional Encoding + Full-Attention) | | 87.73% | 54.51% |
| Informer (Zhou et al., 2021) (Positional Encoding + ProbSparse-Attention) | | 89.26% | 59.40% |
| Autoformer (Wu et al., 2021) (Positional Encoding + Auto-Correlation) | | 89.44% | 63.10% |
| Block-Recurrent Transformer (Hutchins et al., 2022) (Positional Encoding + Full- and Cross-Attention) | | 91.78% | 64.52% |
| THAT (Li et al., 2021a):<br>Temporal Module (GRE + Full-Attention + w/o Recurrent Layer)<br>Channel Module (Positional Encoding + Full-Attention) | | 92.99% | 85.74% |
| **M-GaitFormer: Proposed Transformer** | | | |
| **Temporal Module** | GRE + Full-Attention + w/o Recurrent Layer | 90.96% | 57.06% |
| | GRE + ProbSparse-Attention + w/o Recurrent Layer | 91.07% | 59.48% |
| | GRE + Auto-Correlation + w/o Recurrent Layer | 91.15% | 60.61% |
| | GRE + Auto-Correlation + w/ Recurrent Layer ($N = 8$, $R = 1$, $M = 2$) | 92.23% | 59.20% |
| | **GRE + Auto-Correlation + w/ Recurrent Layer ($N = 9$, $R = 1$, $M = 2$)** | **92.45%** | **68.20%** |
| | GRE + Auto-Correlation + w/ Recurrent Layer ($N = 10$, $R = 1$, $M = 2$) | 91.16% | 53.73% |
| | GRE + Auto-Correlation + w/ Recurrent Layer ($N = 9$, $R = 1$, $M = 1$) | 92.30% | 56.50% |
| | GRE + Auto-Correlation + w/ Recurrent Layer ($N = 9$, $R = 1$, $M = 3$) | 91.10% | 57.06% |
| **Channel Module** | Positional Encoding + Full-Attention | 91.68% | 70.55% |
| | GRE + Full-Attention | 92.28% | 90.77% |
| | GRE + ProbSparse-Attention | 93.26% | 91.20% |
| | **GRE + Auto-Correlation** | **93.64%** | **92.19%** |
| **Temporal + Channel Modules** | **Temporal (GRE + Auto-Correlation + w/ Recurrent Layer)<br>Channel (GRE + Auto-Correlation)** | **94.25%** | **93.33%** |

GRE: Gaussian Range Encoding; $N$, $M$: Number of multi-head Auto-Correlation layers before and after the recurrent layer respectively; $R$: Number of recurrent layers.

and module configuration (Temporal and Channel) are remarked in bold. First, we can see that the Vanilla Transformer outperforms the traditional DL models (CNN, RNN, and CNN + RNN) in both databases. The Vanilla Transformer achieves an accuracy of 87.73% in the whuGAIT database (absolute improvement of 3.19% accuracy compared with the CNN + RNN approach), and 54.51% in the OU-ISIR database (absolute improvement of 7.88% accuracy compared with the CNN + RNN approach). These performance improvements demonstrate the advantages of Transformers compared with traditional CNN and RNN architectures. One example is the ability to train the model using large time sequences, attending to all the previous samples at the same time. In addition, we can also observe a considerable gap in the results between the whuGAIT and OU-ISIR databases. This is due to the OU-ISIR considers a more challenging scenario including many more subjects, sensors, and walking styles. This trend is also observed in the original article for traditional CNN and RNN architectures (Ngo et al., 2014).

The Vanilla Transformer architecture (Vaswani et al., 2017) was improved using ProbSparse-Attention (Informer (Zhou et al., 2021)) and Auto-Correlation (Autoformer (Wu et al., 2021)). Analysing the results included in Table 3.1, we can observe that both Informer and Autoformer outperform the Vanilla Transformer in both whuGAIT and OU-ISIR databases. In particular, for the whuGAIT database, the Informer and Autoformer achieve 89.26% and 89.44% accuracy, respectively, in comparison with the 87.73% accuracy achieved for the Vanilla Transformer (absolute improvement of around 1.60% accuracy). Regarding the OU-ISIR database, much better results are achieved by Informer and Autoformer compared with the Vanilla Transformer (59.40%, 63.10%, and 54.51% accuracy, respectively). Also, Autoformer outperforms Informer in both databases, proving the potential of the multi-head Auto-Correlation mechanism, replacing the point-wise connections for series-wise connections.

The Block-Recurrent Transformer (Hutchins et al., 2022) was presented as an al-

ternative to use the dot-product or periodicity-based series mechanism, which fixes an attention window size. Analysing the results of Table 3.1, the Block-Recurrent Transformer outperforms previous Transformers for both whuGAIT (91.78% accuracy) and OU-ISIR (64.52% accuracy) databases. This is an absolute improvement of 2.34% and 1.42% accuracy compared with Autoformer for the whuGAIT and OU-ISIR databases, respectively.

The THAT Transformer (Li et al., 2021a) proposed a two-stream approach based on Temporal and Channel Modules. This Transformer architecture outperforms all previous Transformers, achieving accuracies of 92.99% and 85.74% for the whuGAIT and OU-ISIR databases, respectively. The improvement is much higher for the OU-ISIR database with an absolute improvement of 21.22% accuracy compared with the Block-Recurrent Transformer. The main reason for this improvement is the proposed GRE in the Temporal Module, better capturing the temporal information of the sample in comparison with the positional encoding considered in all previous Transformers. Moreover, by having multi-scale convolutions instead of feed-forward linear layers, more discriminative patterns of each subject are captured. THAT also demonstrates how, by obtaining features from two points of view (time-over-channel features and channel-over-time features), complementary information can be captured, achieving better performance.

In addition, we show in Table 3.1 the results achieved by our proposed Transformer, M-GaitFormer, under different configurations. First, we analyse the impact in the system performance of each of the modules individually. The Temporal Module with Self-Attention (Full- and ProbSparse-Attention) and without recurrent layer ("w/o recurrent layer" in Table 3.1) achieves values of 90.96% and 91.07% accuracy for the whuGAIT database and 57.06% and 59.48% accuracy for the OU-ISIR database, respectively. These results are further improved by replacing the Self-Attention mechanism with the Auto-Correlation mechanism (91.15% and 60.61% accuracy for the whuGAIT and OU-ISIR

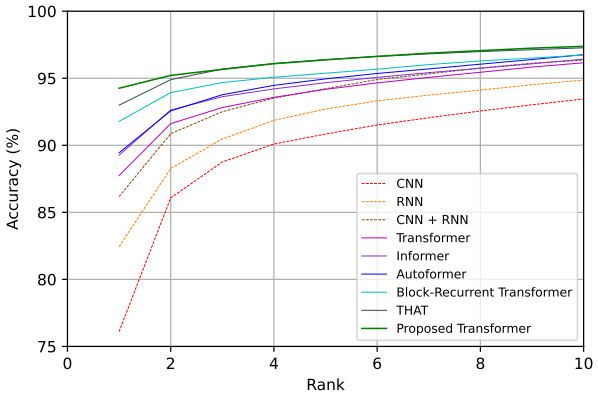
databases, respectively). In addition, when including the recurrent layer ("w/ recurrent layer" in Table 3.1), the Temporal Module achieves better results (92.45% and 68.20% accuracy for the whuGAIT and OU-ISIR databases, respectively), being the best configuration $N = 9$ Multi-head Auto-Correlation layers, $R = 1$ recurrent layer, and $M = 2$ Multi-head Auto-Correlation layers. On the other hand, we can see that the Channel Module with Full-Attention is also able to extract discriminative features for the task, achieving accuracy values of 91.68% and 70.55% for the whuGAIT and OU-ISIR databases, respectively. Moreover, including the GRE (instead of positional encoding), the Channel Module improves the results (92.28% and 90.77% for the whuGAIT and OU-ISIR databases), becoming even better when the Self-Attention mechanism with Full-Attention is replaced by ProbSparse-Attention or Auto-Correlation, (93.26% and 93.64% accuracy for the whuGAIT and 91.20% and 92.19% accuracy for the OU-ISIR databases, respectively).

Finally, we can see how the combination of both Temporal and Channel modules ("Temporal + Channel Modules" in Table 3.1) outperforms all previous Transformer architectures for both whuGAIT (94.25% accuracy) and OU-ISIR (93.33% accuracy) databases. In particular, the proposed Transformer achieves absolute improvements of 2.47% (Block-Recurrent Transformer), 4.81% (Autoformer), 4.99% (Informer), and 6.52% (Vanilla Transformer) accuracy for the whuGAIT database. This improvement is even higher for the OU-ISIR database with absolute improvements of 28.81% (Block-Recurrent Transformer), 30.23% (Autoformer), 33.93% (Informer), and 38.82% (Vanilla Transformer) accuracy. It is important to highlight that in the OU-ISIR database, which is far more challenging than whuGAIT in terms of number of subjects and walking activities, the proposed Transformer achieves considerable improvements in comparison with the THAT approach (93.33% vs. 85.74% accuracy), an absolute improvement of 7.59% accuracy. These results highlight the considerable potential of the proposed

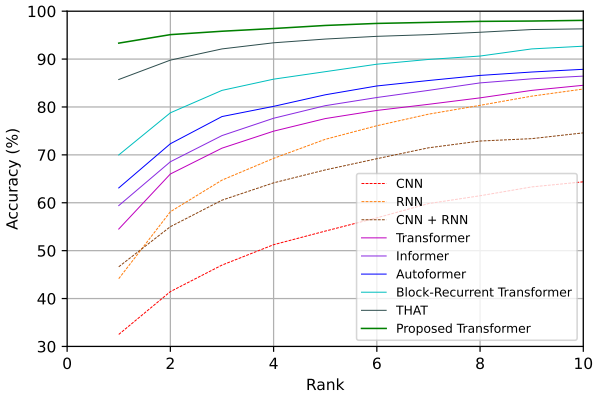
Transformer which are produced for several reasons. First, the GRE allows to introduce in each sample details about its relative position with respect to the contiguous samples (before the Temporal Module) and about the different channels (before the Channel Module), obtaining more complex information. Another advantage is the two-stream architecture, where each of the modules extracts different features (the Temporal Module extracts time features while the Channel Module extracts spatial features). By extracting features from two different perspectives, a more global view of each sample is obtained. In addition, the application of Auto-Correlation in the multi-head Self-Attention mechanism together with the GRE in both Temporal and Channel Modules allow the extraction of series-wise connections in each range of the encoding, analysing the different behaviour of each sample in different environments. Furthermore, including the recurrent layer proposed in the Block-Recurrent Transformer to the Temporal Module offers a comprehensive analysis. The module summarises all the information seen previously, giving a more global view of each sample with respect to the rest. In addition, by including a multi-scale CNN instead of the original feed-forward network, the whole model is series-wise: from the GRE that extracts the position of each sample based on a range of points, multi-head Auto-Correlation with Block-Recurrent Attention, which extracts information periodically based on series, and multi-scale CNN that applies convolutions with different kernels to test the behaviour of samples in different ranges. Finally, the proposed Transformer achieves an absolute improvement of 0.92% in the whuGAIT database (94.25% accuracy) compared with the OU-ISIR database (93.33% accuracy). Some of the differences between the databases that may produce this improvement are: *i)* number of subjects (118 for whuGAIT and 745 for OU-ISIR); *ii)* amount of data available per subject (33,104 training samples for whuGAIT and 13,212 for OU-ISIR); *iii)* different devices (Samsung, Xiaomi, and Huawei smartphones for whuGAIT and three IMUs and a Motorola smartphone for OU-ISIR); and *iv)* different types of walking (walking and non-walking for whuGAIT and walking, slope-up and

**Figure. 3.4:** Cumulative Match Characteristic (CMC) curves of the traditional DL models (CNN, RNN, CNN + RNN) and recent Transformers (Vanilla, Informer, Autoformer, Block-Recurrent, THAT, and the proposed Transformer) for both whuGAIT (a) and OU-ISIR (b) databases. CNNs and RNNs (dashes curves) and Transformer architectures (solid curves).

**(a)** whuGAIT Database




**(b)** OU-ISIR Database




-down for OU-ISIR).

Previous results correspond to the Rank-1 accuracy. Nevertheless, in some applications we might be interested in having a ranked list of possible subjects of interest (e.g., in forensic applications). Fig. 3.4 shows the Cumulative Match Characteristic (CMC) curve of the traditional DL models commonly used in biometric identification (CNN, RNN, CNN + RNN) and recent Transformers (Vanilla, Informer, Autoformer, Block-Recurrent, THAT, and the proposed Transformer, M-GaitFormer) for both whuGAIT and OU-ISIR databases. In general, we can see the same trend in both databases for all approaches, improving the accuracy results with the Rank values. For example, for the proposed Transformer, the accuracy increases from 94.25% (Rank-1) to 97.37% (Rank-10) for the whuGAIT database whereas for the OU-ISIR database this value increases from 93.33% (Rank-1) to 98.08% (Rank-10).

**Table 3.2:** Comparison of the proposed M-GaitFormer with state-of-the-art gait biometric identification approaches in terms of accuracy.

| Study | Method | Database | |
|---|---|---|---|
| | | **whuGAIT** | **OU-ISIR** |
| (Ordóñez and Roggen, 2016) | CNN + RNN | 92.25% | 37.33% |
| (Gadaleta and Rossi, 2018) | CNN + SVM | 92.91% | 44.29% |
| (Zou et al., 2020) | RNN | 91.88% | - |
| | CNN + RNN | 93.52% | - |
| (Tran et al., 2021) | RNN | 93.14% | 78.92% |
| | CNN + RNN | 94.15% | 89.79% |
| **M-GaitFormer** | **Transformer** | **94.25%** | **93.33%** |

### 3.5.2 Experiment 2: Comparison with the State of the Art

Finally, we compare in Table 3.2 the Rank-1 accuracy results achieved by our proposed M-GaitFormer with other state-of-the-art approaches presented in the literature for gait biometric identification: CNNs + SVM (Gadaleta and Rossi, 2018), RNNs (Tran et al., 2021; Zou et al., 2020), and CNNs + RNNs (Ordóñez and Roggen, 2016; Tran et al., 2021; Zou et al., 2020). The best results achieved for each database are remarked in bold. It is important to highlight that all studies consider the same experimental protocol (Zou et al., 2020) for both whuGAIT and OU-ISIR databases, allowing a straightforward and fair comparison between approaches.

In general, our proposed Transformer has outperformed previous approaches in both databases. For the whuGAIT database, the proposed Transformer achieves 94.25% accuracy, showing better results compared with the CNNs + RNNs approach presented in (Tran et al., 2021). Analysing the OU-ISIR database, the proposed Transformer further improves the results achieved by previous approaches with 93.33% accuracy. This is an absolute improvement of 3.54% accuracy compared with the best previous approach (CNNs + RNNs (Tran et al., 2021)). The authors improved the CNN + RNN architecture using an RNN to process each channel, combined in parallel with a CNN

with two channels, one for each sensor. These results support the high potential of the proposed Transformer for gait biometric identification. In addition, it is important to highlight the better time complexity and memory usage of the proposed Transformer compared with traditional DL models.

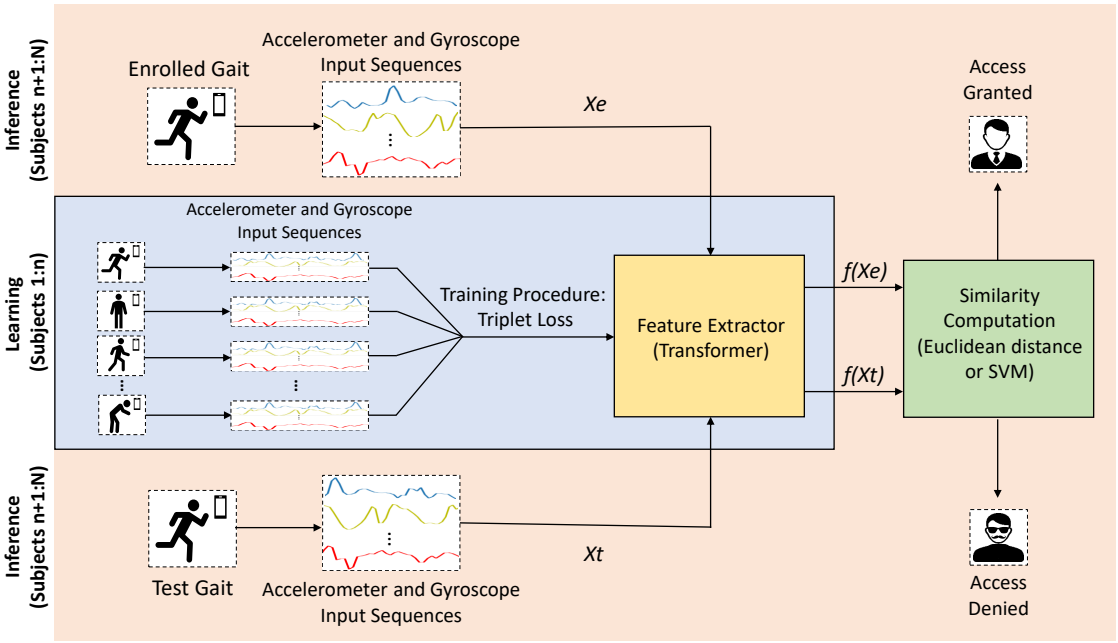
## 3.6 Application of M-GaitFormer to Verification Scenarios

In this Section M-GaitFormer is adapted and evaluated for mobile biometric gait verification. This task has been chosen as it represents a more practical and realistic scenario, such as authorising entries in border control or mitigating fraud risks in financial institutions (Dasgupta et al., 2017).

### 3.6.1 Proposed Approach

Fig. 3.5 provides a graphical representation of the proposed approach, including both learning and inference stages. First, we consider as input the accelerometer and gyroscope time sequences captured by the mobile device. After that, we can observe two main modules: *i)* a feature extractor module based on an adaptation of the proposed Transformer architecture described before for identification scenarios, which is trained on a learning stage using a development dataset, and *ii)* a similarity computation module based on Euclidean distance or SVM, which provides the final verification score of the comparison (inference stage). We describe next each module.

**Figure. 3.5:** Graphical representation of M-GaitFormer, the proposed mobile biometric gait verification system based on Transformers.




$N$: total number of subjects; $X_e$: Enrolment input sequences; $X_t$: Test input sequences; $f(X_e)$: Enrolment feature vector; $f(X_t)$: Test feature vector.

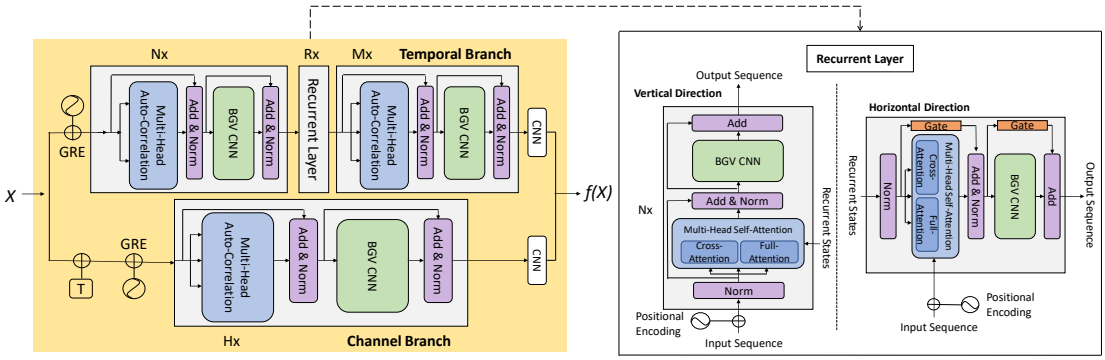
#### 3.6.1.1 Feature Extractor

We consider the same Transformer architecture proposed before in Section 3.2.6 for the feature extraction. For completeness, Fig. 3.6 provides a graphical representation of the Transformer-based feature extractor. This is based on a parallel two-stream architecture with Temporal and Channel Modules.

#### 3.6.1.2 Similarity Computation

As described in Fig. 3.5, the similarity computation module receives as input the features of the enrolled and test gait, $f(X_e)$ and $f(X_t)$, to obtain the final similarity score. Three

**Figure. 3.6:** Graphical representation of the Transformer-based Feature Extractor.




$X$: Input sequences; $f(X)$: Feature vector; GRE: Gaussian Range Encoding; T: Transposition; Nx, Rx, Mx, Hx: Number of layers of each type; BGV CNN: Biometric Gait Verification CNN.

different configurations are studied:

- **Euclidean Distance** is a simple but very popular approach in biometrics based on the distance between the feature vectors $f(X_e)$ and $f(X_t)$:

$$d(X_e, X_t) = \|f(X_e) - f(X_t)\| \tag{3.9}$$

- **One-Class Support Vector Machine (OC-SVM)** comprises of training a single specific SVM classifier per subject. In this particular configuration (one-class), only the enrolment samples of the subject are considered to train the SVM.

- **Binary Support Vector Machine (B-SVM)** is similar to the OC-SVM in training one specific SVM classifier per subject. The main difference is that for each subject, one classifier is trained using both enrolment samples of the subject and also gait samples of other subjects (from a development dataset), acting as impostors.

### 3.6.2 Experimental Setup

This Section provides the details of the experimental framework of the study. First, we describe in Section 3.6.3 the system configuration of the proposed M-GaitFormer. Then, Section 3.6.4 presents the standard experimental protocol considered for whuGAIT and OU-ISIR databases for verification scenarios.

### 3.6.3 System Details

Regarding the Transformer-based feature extractor, the GRE contains $G = 20$ Gaussian distributions. The Temporal Branch contains $N = 9$, $R = 1$, and $M = 2$ layers, and $F = 8$ independent heads for each layer whereas the Channel Branch comprises $H = 1$ layer and $F = 6$ independent heads for each layer. In both branches, the BGV CNN contains 3 convolutional layers with $L$ units each, ReLU activation functions, and kernel sizes 1, 3, and 5 respectively, followed by dropout layers with a rate of 0.1. Finally, after the Temporal and Channel Branches we consider 2 convolutional layers with $L$ units each, ReLU activation functions, and kernel sizes of 512 and 256 respectively.

The Transformer-based feature extractor is trained with a triplet loss function, using Euclidean distance with a margin $\alpha = 1.0$. Adam optimiser is considered with a learning rate of 0.001. It is trained using a stop condition: if the feature extractor does not achieve better results in the validation dataset during 15 epochs, the training stops. Regarding the similarity computation module, the OC-SVM has an *RBF* kernel and $\gamma = 1.0$ while B-SVM has an *RBF* kernel and $\gamma = 0.5$. Experiments are implemented in `PyTorch`.

### 3.6.4 Experimental Protocol

We describe next the details of the experimental protocol considered for each database and stage (learning and inference):

- WhuGAIT: We follow the experimental protocol proposed in (Tran and Choi, 2020). From the 118 total subjects, 98 are used in the learning stage for training the feature extractor while the remaining 20 unseen subjects are only considered for the final evaluation (inference stage). Regarding the learning stage, we build triplets using the 98 subjects of the development dataset. Each triplet comprises two genuine samples of the same subject (enrolment and genuine test), and a third one from a different subject (impostor test). The genuine and impostor test samples included in the triplets are selected randomly with a uniform distribution. Considering all possible triplets, a total of 284,030 triplets are included for the feature extractor training.

- OU-ISIR: We follow the same experimental protocol presented in (Fernandez-Lopez et al., 2019; Tran et al., 2021). From the 744 total subjects, 520 are used in the learning stage for training the feature extractor while the remaining 224 unseen subjects are part of the final evaluation (inference stage). Concerning the learning stage, we build triplets using the 520 subjects of the development dataset, following the same approach described for the WhuGAIT database. A total of 229,543 triplets are considered for training the feature extractor.

Regarding the inference stage, we consider in both WhuGAIT and OU-ISIR databases the same experimental protocol. For each unseen subject of the evaluation dataset, we follow the same experimental protocol presented in (Fernandez-Lopez et al., 2019; Tran and Choi, 2020; Tran et al., 2021), i.e., 50% of the samples of the subject selected ran-

domly are used as enrolment, while the remaining 50% are considered for testing in order to obtain the genuine scores. Impostor scores are obtained comparing the enrolment samples of the subject with samples of the remaining subjects of the evaluation dataset (same number of genuine and impostor comparisons). Depending on the similarity computation approach considered (i.e., Euclidean distance, OC-SVM, B-SVM), the final score is calculated differently. For the Euclidean distance, the final score is the average of the scores obtained when comparing one test sample (genuine/impostor) with each enrolment sample. For the SVM approaches, the final score is obtained when comparing one test sample (genuine/impostor) with the specific SVM model created with all enrolment samples of that subject.

### 3.6.5 Experimental Results

Section 3.6.5.1 provides an analysis of the proposed M-GaitFormer and each of its modules, for both whuGAIT and OU-ISIR databases, and also for the different similarity computation configurations (i.e., Euclidean distance, OC-SVM, and B-SVM). Finally, we compare in Section 3.6.5.2 our proposed M-GaitFormer with the state of the art using the same experimental protocol.

#### 3.6.5.1 Experiment 1: M-GaitFormer Results

Table 3.3 shows the results of our proposed M-GaitFormer in terms of EER (%) for the whuGAIT and OU-ISIR evaluation datasets and for the different similarity computation configurations considered: Euclidean distance, OC-SVM, and B-SVM. In addition, we include: *i)* the results achieved by the Vanilla Transformer (Vaswani et al., 2017), and *ii)* the contributions in the performance of each of the branches considered in M-GaitFormer.

**Table 3.3:** Results of our proposed M-GaitFormer in terms of EER (%) for the whuGAIT and OU-ISIR evaluation datasets and for the different similarity computation configurations considered: Euclidean distance, OC-SVM, and B-SVM. In addition, for completeness, we include: *i)* the results achieved by the Vanilla Transformer (Vaswani et al., 2017), and *ii)* the contributions in the performance of each of the branches considered in M-GaitFormer.

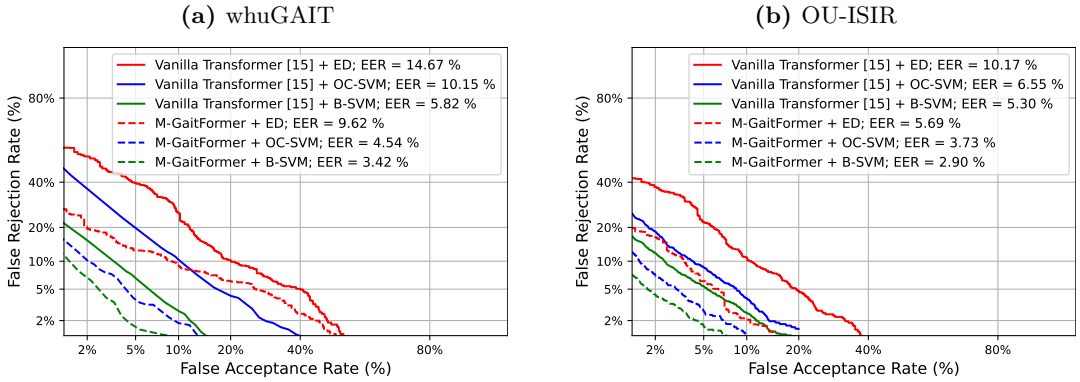
| **Method** | | **Databases** | |
|---|---|---|---|
| **Feature Extractor** | **Similarity Computation** | **whuGAIT** | **OU-ISIR** |
| Vanilla Transformer (Vaswani et al., 2017) | Euclidean distance | 14.67 | 10.17 |
| | OC-SVM | 10.15 | 6.55 |
| | B-SVM | 5.82 | 5.30 |
| M-GaitFormer (Temporal Branch w/o Recurrent Layer) | Euclidean distance | 13.03 | 9.15 |
| | OC-SVM | 8.02 | 6.89 |
| | B-SVM | 4.02 | 5.13 |
| M-GaitFormer (Temporal Branch w/ Recurrent Layer) | Euclidean distance | 11.97 | 8.59 |
| | OC-SVM | 7.70 | 6.78 |
| | B-SVM | 4.11 | 4.79 |
| M-GaitFormer (Channel Branch) | Euclidean distance | 13.97 | 9.80 |
| | OC-SVM | 7.38 | 7.05 |
| | B-SVM | 4.25 | 4.15 |
| **M-GaitFormer (Temporal + Channel Branches)** | Euclidean distance | 9.62 | 5.69 |
| | OC-SVM | 4.54 | 3.73 |
| | **B-SVM** | **3.42** | **2.90** |

First, we analyse the impact in the system performance of each of the branches considered in the proposed M-GaitFormer. To provide a better understanding of the results, we focus now on the Euclidean distance configuration, as the proposed Transformer is used as feature extractor. The Temporal Branch (without the recurrent layer) achieves values of 13.03% and 9.15% EER for the whuGAIT and OU-ISIR databases, respectively. These results are further improved if we include the recurrent layer in the Temporal Branch, i.e., 11.97% and 8.59% EER for the whuGAIT and OU-ISIR databases, respectively. In addition, we can see that the Channel Branch is also able to extract discriminative features for the task, achieving EER values of 13.97% and 9.80% for the whuGAIT and OU-ISIR databases. Finally, we can see in Table 3.3 how the combination of both Temporal and Channel Branches achieves the best results in both whuGAIT (9.62% EER) and OU-ISIR databases (5.69% EER). These results prove the potential of our proposed Temporal and Channel Branches for the feature extraction.

Analysing the impact of the similarity computation configuration, we can see that in general, the Euclidean distance provides worse results compared to the case of training classifiers such as SVM. For example, focusing on the M-GaitFormer (Temporal + Channel Branches) in Table 3.3, the Euclidean distance achieves values of 9.62% and 5.69% EER for the whuGAIT and OU-ISIR databases, respectively. These results are further improved when considering the B-SVM (3.42% and 2.90% EER, respectively), with relative improvements of 64.45% and 49.03% EER. These results evidence the importance of using classifiers such as SVM to better adapt the features extracted by the Transformer to each specific subject. This is in accordance with related works that have shown subject-adaptation (Fierrez et al., 2018a) to be very useful in behavioural biometrics (Fierrez et al., 2005).

Finally, we compare in Table 3.3 the results achieved by our proposed M-GaitFormer (Temporal + Channel Branches) with the original Vanilla Transformer (Vaswani et al.,

**Figure. 3.7:** DET curves and EER (%) results on the (a) whuGAIT and (b) OU-ISIR evaluation datasets for the Vanilla Transformer (Vaswani et al., 2017) and the proposed M-GaitFormer in the three similarity computation configurations considered: *i)* Euclidean distance (ED), *ii)* One-Class SVM (OC-SVM), and *iii)* Binary SVM (B-SVM). Vanilla Transformer (solid curves) and M-GaitPrivacyON (dashes curves).




2017). In addition, for completeness, we include in Fig. 3.7 the Detection Error Trade-Off (DET) curves of the proposed M-GaitFormer and the Vanilla Transformer (Vaswani et al., 2017) for both whuGAIT and OU-ISIR databases. In general, we can observe that M-GaitFormer outperforms the Vanilla Transformer in all configurations (Euclidean distance, OC-SVM, and B-SVM), proving the potential of the proposed method. For the Euclidean distance configuration, M-GaitFormer achieves relative improvements of 34.42% and 44.05% EER for the whuGAIT and OU-ISIR databases whereas for the B-SVM configuration the relative improvements are 41.24% and 45.28% EER, respectively. Finally, the best results achieved by the proposed M-GaitFormer are 3.42% and 2.90% EER for the whuGAIT and OU-ISIR databases.

#### 3.6.5.2 Experiment 2: Comparison with the State of the Art

Table 3.4 provides a comparison in terms of EER (%) of the proposed M-GaitFormer with others state-of-the-art approaches in the literature for both whuGAIT and OU-ISIR evaluation datasets. Note that in some cases, indicated in Table 3.4 with the symbol

**Table 3.4:** Comparison of the proposed M-GaitFormer system with state-of-the-art approaches in mobile biometric gait verification in terms of EER (%) for the whuGait and OU-ISIR evaluation datasets (Ngo et al., 2014; Tran and Choi, 2020). Note that the symbol * indicates those studies that do not use the standard experimental setup considered in the literature.

| Study | Method | Database | |
|---|---|---|---|
| | | **whuGAIT** | **OU-ISIR** |
| (Nguyen et al., 2017) | CNN | - | 10.43 |
| (Fernandez-Lopez et al., 2019)* | LSTM | - | 7.55 |
| (Subramanian and Sarkar, 2018)* | Kabsch alignment | - | $> 6.00$ |
| (Tran and Choi, 2020) | CNN | - | 4.49 |
| (Zou et al., 2020) | LSTM | 7.50 | - |
| | CNN & LSTM | 6.50 | - |
| (Tran et al., 2021)* | LSTM | 5.82 | 6.63 |
| | CNN & LSTM | 4.52 | 3.36 |
| **M-GaitFormer** | **Transformer** | **3.42** | **2.90** |

*, the studies do not use the standard experimental setup considered in the literature, therefore the results must be interpreted carefully. Despite of that, it is patent that the proposed M-GaitFormer achieves state-of-the-art results in both whuGAIT and OU-ISIR databases.

Analysing the results on the whuGAIT database, our proposed M-GaitFormer achieves an EER of 3.42%, a relative EER improvement of 54.40% and 41.24% compared with systems based on LSTM architectures (Tran et al., 2021; Zou et al., 2020). Furthermore, M-GaitFormer outperforms previous approaches in the literature based on CNN & LSTM (Tran et al., 2021; Zou et al., 2020) with relative improvements of 47.39% and 24.34% EER.

A similar trend can be observed for the OU-ISIR database. Our M-GaitFormer model achieves a relative EER improvements of 72.20% and 35.41% compared to traditional CNN architectures (Nguyen et al., 2017; Tran and Choi, 2020). In addition, M-GaitFormer achieves a relative EER improvement of 61.59% and 56.26% in comparison with LSTM architectures (Fernandez-Lopez et al., 2019; Tran et al., 2021). Finally, M-GaitFormer reaches a relative improvement of 13.69% EER in comparison with the

CNN & LSTM architecture presented in (Tran et al., 2021).

This comparison with the state of the art proves the potential of our proposed M-GaitFormer architecture for the task of mobile biometric gait verification, outperforming previous approaches based on CNN, LSTM, or a combination. Some of the advances that have been achieved by applying Transformers are: *i)* application of Auto-Correlation and attention mechanisms, which allow operating over long sequences; *ii)* operation on all previous samples of the time sequence at the same time, without the need to summarise; *iii)* extraction of features from two different perspectives (Channel and Temporal Branches), obtaining more discriminative information; and *iv)* inclusion of a GRE together with the Auto-Correlation and the Block-Recurrent attention to extract features over the entire time sequence instead of single points, considering important aspects in time sequences such as the samples distribution over the time.

## 3.7 Conclusions

This Chapter has explored and proposed novel behavioural biometric systems based on Transformers. To the best of our knowledge, this is the first study that presents a complete framework for the use of Transformers in gait biometrics. Several Transformer architectures (Vanilla, Informer, Autoformer, Block-Recurrent Transformer, and THAT) are considered in the experimental framework, together with a new proposed configuration, M-GaitFormer. Two popular public databases are considered in the analysis, whuGAIT and OU-ISIR.

For the identification task, the proposed M-GaitFormer has outperformed previous Transformer architectures and traditional DL architectures (i.e., CNNs, RNNs, and CNNs + RNNs) in both databases. In particular, for the challenging OU-ISIR database,

M-GaitFormer achieves 93.33% accuracy, resulting in accuracy absolute improvements compared with other techniques of 7.59% (THAT), 28.81% (Block-Recurrent Transformer), 30.23% (Autoformer), 33.93% (Informer), and 38.82% (Vanilla Transformer). M-GaitFormer has also been compared with state-of-the-art gait recognition systems, outperforming them. In addition, it is important to highlight the enhanced time complexity and memory usage of M-GaitFormer compared with traditional CNN and RNN models.

For the verification task, our proposed M-GaitFormer follows the same trend, outperforming the state of the art with results of 3.42% and 2.92% EER on whuGAIT and OU-ISIR databases, respectively.

Finally, we have concluded the Chapter by clarifying the reasons why M-GaitFormer might outperform previuos approaches presented the literature on both identification and verification tasks.

# Chapter 4

# SwipeFormer: Mobile Biometric Swipe Authentication

## 4.1 Introduction

In this Chapter we investigate the potential application of Transformers to another behavioural biometric trait: touchscreen biometrics. We present a novel verification system designed to study unconstrained (commonly referred to as "in-the-wild") swipe gestures in a free-direction environment. This approach sets itself apart from popular touchscreen biometric systems that solely consider swipe gestures in specific directions, such as horizontal or vertical.

This Chapter is structured as follows: Section 4.2 describes the architecture of SwipeFormer, including the Transformer-based feature extractor and the different similarity computation approaches. Section 4.3 describes the databases considered while Section 4.4 provides a description of the system details and experimental protocol. Sec-

tion 4.5 describes the results achieved and comparison with the state of the art. For completeness, we also show in Section 4.6 the evaluation of these Transformer architectures on a keystroke free-text verification scenario. Finally, Section 4.7 draws the final conclusions. This Chapter is based on the following publications: (Delgado-Santos et al., 2023b; Stragapede et al., 2023a,b).

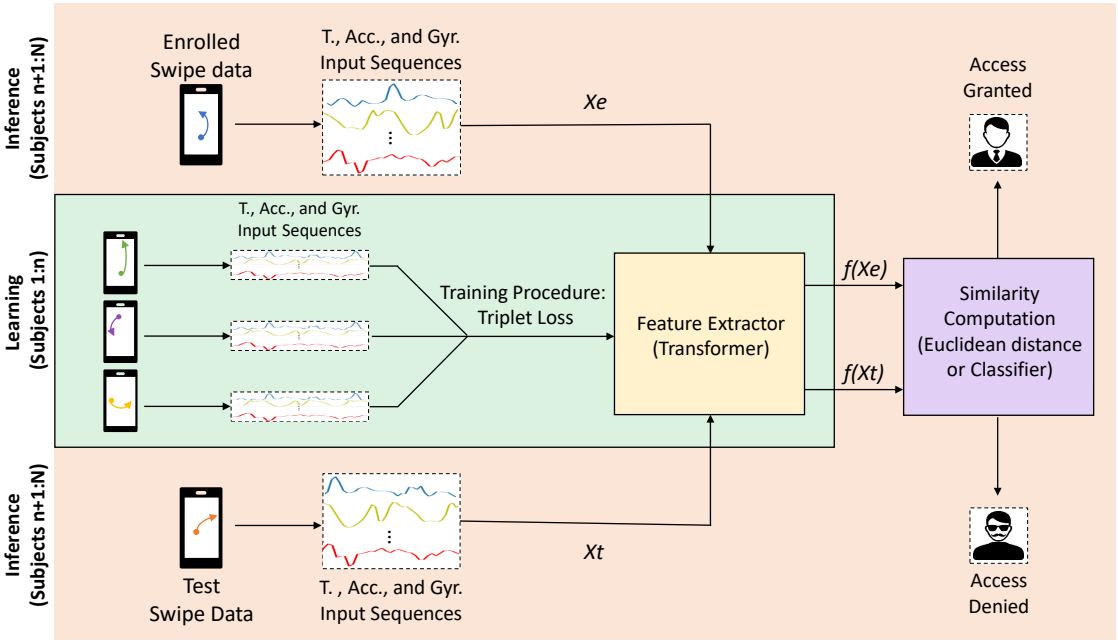
## 4.2 Proposed Approach: SwipeFormer

Fig. 4.1 provides a general representation of SwipeFormer, our proposed touchscreen verification system for mobile scenarios. As can be seen, SwipeFormer shares some similarities in comparison with M-GaitFormer presented before in Chapter 3. The previous architecture has been adapted to this new biometric trait in order to extract more discriminative features. First, time sequences from the touchscreen and background sensors (i.e., accelerometer and gyroscope) are captured and introduced as input of the system. Subsequently, SwipeFormer consists of two modules, similar to M-GaitFormer for the verification case: *i)* a feature extractor based on a Transformer architecture, trained in the learning stage with a development dataset; and *ii)* a similarity computation module, which provides the final similarity scores using an evaluation dataset based on subjects not seen in the learning stage (inference stage). The specific details of each module are described next.

### 4.2.1 Feature Extractor

Fig. 4.2 shows a graphical representation of the Transformer-based feature extractor trained in the learning stage, based on an adaptation of the architecture presented in Section 3.7. SwipeFormer comprises two modules in parallel: *i)* a Temporal Module,

**Figure. 4.1:** Graphical representation of SwipeFormer, the proposed mobile touchscreen biometric verification system based on Transformers.

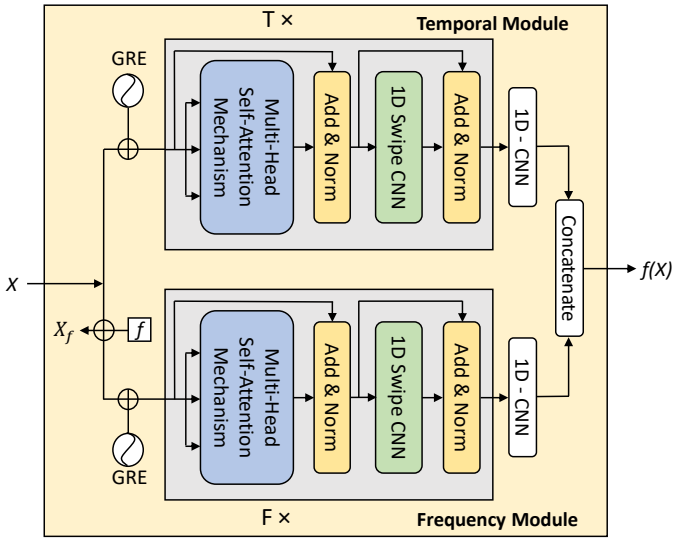



$N$- total number of subjects; $X_e$- Enrolled swipe sequences; $X_t$- Test swipe sequences; $f(X_e)$- Enrolled feature vector; $f(X_t)$- Test feature vector; T.- Touch; Acc.- Accelerometer; Gyr.- Gyroscope.

inspired by M-GaitFormer, and *ii)* a novel Frequency Module, responsible for extracting discriminative features in the frequency domain. Although in (Zhang et al., 2022b) the authors demonstrate that models of attention in various domains (i.e., temporal and frequency) are considered equivalent when exposed to linear conditions, the study also demonstrates the various behaviours exhibited in different domains.

Analysing the Temporal Module first, the input swipe sequence $X$ with $L$ time samples is shaped by a GRE, similar to M-GaitFormer. Following the GRE, the Temporal Module contains a sequential stack of $T$ layers. Each layer contains two sub-layers: *i)* a multi-head Self-Attention mechanism, and *ii)* a one-dimensional multi-scale swipe CNN created specifically for this task.

**Figure. 4.2:** Graphical representation of the Transformer-based Feature Extractor.




$X$- Input swipe sequence; $X_f$- Input swipe sequence in the frequency domain; $f(X)$- Feature vector; GRE- Gaussian Range Encoding; f- Frequency transformation; T ×, F ×- Number of layers of each type; 1D-CNN- One Dimension Convolutional Neural Network.

Considering the Frequency Module, the input swipe sequence $X$ is represented in the frequency domain by a discrete Fourier transformation $X_f$. After this, a GRE is included to preserve frequency information. Following an identical architecture to the Temporal Module, the Frequency Module contains the same two sub-layers: *i)* a multi-head Self-Attention mechanism, and *ii)* a one-dimensional multi-scale swipe CNN. Each sub-layer is also followed by a residual connection and a layer normalisation (Add & Norm in Fig. 4.2).

After the Time and Frequency Modules, a one-dimensional convolutional block is included similar to M-GaitFormer. The features extracted by each module are concatenated and fed into a dense layer with sigmoid activation, obtaining the output vector $f(X)$:

$$f(X) = [\text{CNN}(f_t(X)); \text{CNN}(f_f(X))] \tag{4.1}$$

where $f_t(X)$ and $f_f(X)$ are the extracted features from the Temporal and Frequency Modules respectively.

### 4.2.2 Similarity Computation

The feature vectors extracted from the enrolled $f(X_e)$ and test $f(X_t)$ swipe sequences are introduced in the similarity computation module to obtain the final similarity score as described in Fig. 4.1. Six different approaches are considered at inference stage:

- **Euclidean Distance** is a popular and simple approach widely used in biometrics as it does not require any training. It simple compares the similarity between feature vectors based on the subtraction. Euclidean distance calculates the distance between $f(X_e)$ and $f(X_t)$:

$$d(X_e, X_t) = \sqrt{(f(X_e) - f(X_t))^2} \tag{4.2}$$

- **Shrunk Covariance** reduces the ratio between the smallest and the largest eigenvalues of the empirical covariance matrix finding the l2-penalised Maximum Likelihood Estimator of the covariance matrix. This matrix is commonly used to model the statistical relationships among the features extracted from biometric samples. The Shrunk covariance is fitted for each subject and tested with samples from the same subject (genuine), and from other subjects in the final evaluation dataset (impostor).

- **Kernel Density Estimator (KDE)** applies kernel smoothing for probability

density estimation. KDE is a flexible and powerful tool for analysing and modelling biometric trait distributions, enabling a good understanding of data. Each estimator is trained on data from a single subject and tested on genuine and impostor samples.

- **Gaussian Mixture Model (GMM)** shapes the feature vector from a subject into a series of Gaussians in a probabilistic way with Expectation-Maximization (EM) algorithm. GMM is well-known method in biometrics as it can represent complex, multi-modal distributions, capture intra-class variability, and reduce dimensionality. For the final evaluation, the model is tested with samples from the same subject (genuine), and from other subjects in the final evaluation dataset (impostor).

- **One-Class Support Vector Machine (OC-SVM)** is trained per subject. This configuration may be suitable for many application scenarios as it only considers data from the genuine subject, mapping the data into a high-dimensional feature space where a linear decision boundary can effectively separate the target class from outliers. In the one-class configuration only the enrolled samples of the subject are considered to train the SVM.

- **Binary Support Vector Machine (B-SVM)** trains a subject-specific SVM classifier. In contract to OC-SVM, one classifier is trained using both enrolled samples of the subject and also samples of other subjects (from the development dataset) used as impostor. In cases where genuine and impostor samples are available for training, it is a very powerful classifier as it provides strong generalisation, is robust to overfitting and there are few parameters to adjust.

**Table 4.1:** Summary of the main characteristics of the databases considered in this study together with their experimental setup. T.-Touch, Acc.- Accelerometer, Gyr.- Gyroscope; x- x axis; y- axis; p- pressure.

| Database | Subjects | Device | Sessions | Features | Length Swipes |
|---|---|---|---|---|---|
| In-House | Android: 232<br>iOS: 232 | Free<br>(Android & iOS) | 2<br>($\geq$ 1 week) | T. (x, y, t, area)<br>Acc. (x, y, z)<br>Gyr. (x, y, z) | Android: 30<br>iOS: 10 |
| (Frank et al., 2012) | 41 | HTC Droid Inc.<br>Google Nexus One<br>Google Nexus S<br>Samsung Galaxy S | 2<br>($\geq$ 1 week) | T. (x, y, p, t, area) | 50 |
| (Acien et al., 2021a) | 600 | Free<br>(Android) | $\leq$ 5<br>($\geq$ 1 day) | T. (x, y, p, t, area)<br>Acc. (x, y, z)<br>Gyr. (x, y, z)<br>... | 100 |

## 4.3 Databases Description

Three different databases have been considered in the experimental framework of this Chapter. These databases contain touchscreen data extracted from mobile devices while performing swipe gestures. In particular, we consider an in-house collected database together with two public databases widely used in the literature. In each database, different acquisition conditions and mobiles devices are considered (e.g., sampling rate, screen size, etc.). The main characteristics of the databases together with the experimental setup followed in this Chapter are reported in Table 4.1.

### 4.3.1 In-House Database

This database comprises 464 subjects. For each swipe gesture acquired, we have the corresponding information of the touchscreen, accelerometer, and gyroscope sensors. The subjects were required to authenticate themselves in real-world, unconstrained (not supervised) settings. This scenario allowed subjects to use their personal devices freely, performing the gestures in their preferred way, location, and timing. In addition, there were no restrictions in terms of the position of the device (portrait or landscape) and

the direction of the gestures (vertical or horizontal). As a result, this database considers real operational conditions, unlike previous swipe databases in the field. The data were collected using each subject's personal smartphone (Android and iOS) over a period of one year, with a minimum one-week gap between at least two sessions. To comply with the General Data Protection Regulation (GDPR) from the European Commission, information related to the device specifications (only the operating system), demographics, and statistics were not collected.

### 4.3.2 Frank Database

The Frank database (Frank et al., 2012) contains swipe data from 41 subjects. Touchscreen data were collected from 4 different Android devices (HTC Droid Inc, Google Nexus One, Google Nexus S, and Samsung Galaxy S) while subjects were completing an image comparison task and reading text under constrained conditions. All devices are in portrait orientation. At least 2 sessions per subject separated by 1 week were collected.

### 4.3.3 HuMI Database

The HuMI database(Acien et al., 2021a) is the largest public mobile touchscreen database available in the literature. Data were collected from 600 subjects interacting with a touchscreen alongside different background sensors (e.g., linear accelerometer, accelerometer, and gyroscope, among others). Data were collected using an Android app while subjects performed 8 simple tasks (i.e., keystroke, swipe, tap, audio, and draw a number) on their own devices under constrained conditions. The number of acquisition sessions per subject is 5 or fewer, with at least 1 day in between.

**Table 4.2:** Hyperparameters configuration.

| | |
|---|---|
| Temporal Module | Gaussians in GRE (G) = 20 |
| | Temporal Layers (T) = 9 |
| | Temporal Heads (H) = 20 |
| Frequency Module | Gaussians in GRE (G) = 20 |
| | Frequency Layers (F) = 9 |
| | Frequency Heads (H) = 20 |
| Temporal + Frequency Modules | Feature Vector Size (S) = 64 |

## 4.4 Experimental Setup

### 4.4.1 Feature Extractor Hyperparameters

This Section describes the hyperparameters of the proposed Transformer. The best architecture and hyperparameters of the proposed Transformer have been selected using only the development dataset of the in-house database. This selection has been carried out manually, based on trial and error. Table 4.2 provides an overview of the key hyperparameters of SwipeFormer. Both Temporal and Frequency modules have the same structure, the only difference is the FFT included in the Frequency Module with an output size of $s = L - 1$, where $L$ is the length (number of samples) of each input swipe sequence. The GRE includes $G = 20$ Gaussian distributions. After these, the Temporal Module comprises $T = 9$ layers, and the Frequency Module contains $F = 9$ layers. Each layer includes $H = 10$ heads. Subsequently, in each module the multi-scale swipe CNN comprises 3 convolutional layers with $L$ units each, and kernel sizes 1, 3, and 5 respectively. In addition, the convolutional layers include ReLU activation functions, followed by dropout layers with a rate of 0.1. Finally, 2 convolutional layers with $L$ units each, ReLU activation functions, and kernel sizes of 512 and 256 are included at the end of each module. The final output vector $f(X)$ contains $S = 64$ features as a result of concatenating the output of the modules fed into the dense layer with a sigmoid activation.

### 4.4.2 System Details

The Transformer-based feature extractor is trained in the learning stage. The triplet loss strategy is employed for the training, including a Euclidean distance with a margin of $\alpha = 1.0$ for each triplet comparison. Each triplet consists of three swipe gestures (containing each swipe the corresponding information related to the touchscreen, accelerometer, and gyroscope sensors): *i)* anchor (belonging to an enrolled subject), *ii)* positive (belonging to the same subject considered in the anchor), and *iii)* negative (belonging to a different subject). Triplets are randomly formed using the subjects and swipes gestures included in the training dataset, following the guidelines explained before for the anchor, positive, and negative samples. The Adam optimiser with a learning rate of 0.001 is used. Furthermore, a stop condition is included for the training: if the feature extractor does not improve the validation loss for 10 epochs, the training stops.

In the inference stage, the evaluation of SwipeFormer includes different similarity computation approaches. The Shrunk Covariance and KDE are evaluated with the Mahalanobis distance; and GMM with diagonal covariance. In addition, KDE uses a Guassian kernel and a bandwidth of 0.9. OC-SVM and B-SVM contain an *RBF* kernel with $\gamma = 0.5$. SwipeFormer is implemented in `PyTorch`.

### 4.4.3 Experimental Protocol

Next, we describe the experimental protocol details considered in this Chapter. The specifications of each database and stage (learning and inference) are included.

#### 4.4.3.1 Experiment 1: In-House Database

The first experiment analyses the performance of SwipeFormer using the in-house database. Data from the touchscreen (x, y, area pressed in the screen), accelerometer (x, y, z), and gyroscope (x, y, z) are included. Two experiments are considered, one for Android and one for iOS operating systems, both employing the same experimental protocol.

In each experiment, the learning stage consists of 190 subjects, 148 of which belong to the train dataset and 42 to the validation dataset. Regarding the inference stage, the 42 remaining unseen subjects are included in the evaluation dataset. Regarding the learning stage, the training dataset contains in total 18,066 triplets and the validation dataset includes 3,778. Each swipe of each subject represents the anchor of a triplet, the positive pair is randomly selected from another swipe of the same subject, and the negative pair from a random swipe of another subject of the training/validation dataset. Regarding the inference stage, for each subject the final verification scores are obtained comparing 5 enrolled swipes from the first session with 10 test swipes from the last session (genuines) and with 10 swipe from other subjects (impostors). Finally, the Android subset contains a sequence length $L = 30$ while the iOS subset $L = 10$ due to the sample frequency.

#### 4.4.3.2 Experiment 2: Frank and HuMI Databases

To validate the potential of SwipeFormer, the proposed architecture has been evaluated and compared with the literature using two popular publicly available databases: *i)* the database proposed in (Frank et al., 2012), which is one of the first public databases in the literature, and *ii)* HumiDB proposed in (Acien et al., 2021a), which is the largest public database in the field, acquired recently in unconstrained scenarios.

First, in the Frank DB, 33 subjects are included in the development dataset (learning stage) while the remaining 8 subjects are included for the final evaluation (inference stage). Data from the touchscreen (x, y, area) are included. For the learning stage, the training dataset includes 11,694 triplets and the validation dataset contains 3,365 triplets. All swipes have $L = 50$ samples. In order to provide a fair comparison with the literature, we follow the same (inter-session) experimental protocol considered in (Fierrez et al., 2018b) for the inference stage. It is important to highlight that, unlike previous approaches in the literature that consider intra-session variability (Fierrez et al., 2018b; Zhang et al., 2015), we follow an inter-session experimental protocol where enrolment and test samples are from different sessions in time (different days), being a more realistic and challenging scenario for behavioural biometrics. For each subject, one session is used for enrolment while the other one is used for test as genuine. Regarding impostor scores, swipes from other random subjects are used as impostors. It is important to highlight that, contrary to previous approaches in the literature such as (Fierrez et al., 2018b), SwipeFormer considers a more universal scenario, without specifying the particular swipe directions (vertical, horizontal, etc.) or position of the device (portrait or landscape). Therefore, we consider in the analysis more challenging scenarios.

In addition, for HuMIDB, we replicate the experimental protocol considered in (Acien et al., 2020). In particular, the right-swipe gestures between tasks are included in this study. Data from the touchscreen (x, y, pressure) are considered. Specifically, 424 subjects are used in the learning stage (24,430 triplets for training and 5,946 triplets for validation), while the remaining 178 unseen subjects are part of the final evaluation (inference stage). All swipes have a length of $L = 100$ samples. Finally, for the inference stage, we consider the first 5 swipes per subject as enrolled swipes, and the last 10 genuine swipes for testing. Furthermore, 10 random swipes from other subjects are included as impostor swipes for testing.

**Table 4.3:** Comparison of the performance in EER (%) achieved by the proposed SwipeFormer with different similarity computation approaches in our in-house database (Android and iOS devices).

| **Method** | | **Databases** | | | |
|---|---|---|---|---|---|
| **Feature Extractor** | **Similarity Computation** | **Android** | | **iOS** | |
| | | **T.** | **T., Acc., Gyr.** | **T.** | **T., Acc., Gyr.** |
| **SwipeFormer** | Eucl. Dist. | 12.30 | 12.10 | 13.90 | 13.70 |
| | Shrunk Cov. | 13.00 | 11.90 | 9.00 | 11.20 |
| | KDE | 7.80 | 7.50 | 7.90 | 8.50 |
| | GMM | 10.40 | 10.80 | 8.60 | 8.10 |
| | OC-SVM | 8.70 | 7.60 | 8.30 | 9.90 |
| | **B-SVM** | 6.90 | **6.60** | 5.30 | **3.60** |
| (Fierrez et al., 2018b) | | 43.40 | 43.20 | 35.10 | 34.80 |
| (Acien et al., 2020) | | 18.10 | 17.70 | 15.30 | 14.70 |

Eucl. Dist.- Euclidean Distance; Shrunk Cov.- Shrunk Covariance; T.- Touch; Acc.- Accelerometer; Gyr- Gyroscope.

## 4.5 Experimental Results

### 4.5.1 Experiment 1: In-House Database

Table 4.3 shows the results of SwipeFormer in terms of verification EER (%) for the Android and iOS evaluation datasets and for the different similarity computation configurations considered: Euclidean distance, Shrunk Covariance, KDE, GMM, OC-SVM, and B-SVM. Two different feature configurations are studied: *i)* including the touchscreen and *ii)* the combination of touchscreen and background sensors (accelerometer and gyroscope). In addition, to provide a better comparison of SwipeFormer with the state of the art, we include in the table the results achieved by recent approaches in the literature, i.e., (Fierrez et al., 2018b) and (Acien et al., 2020).

Analysing the results with Euclidean distance, SwipeFormer achieves EER values of 12.30% for touchscreen and 12.10% for touchscreen and background sensors (relative EER improvement of 1.70%) in the Android configuration; and 13.90% for touchscreen and 13.70% for touchscreen and background sensors (relative EER improvement

of 1.40%) in the iOS configuration. These results demonstrate how background sensors on mobile devices can provide additional information (e.g., the uniqueness of the device used and held by the subject).

In addition, we analyse the performance of different similarity computation configurations with the best feature configuration (touchscreen and background sensors) in the Android configuration. Compared with Euclidean distance, other approaches improve this result in terms of ERR with relative improvements of 1.70%, 38.00%, 10.70%, 37.20%, 45.50% for Shrunk Covariance, KDE, GMM, OC-SVM, and B-SVM respectively. These results prove that by training each SVM with both genuine (from the enrolled subject) and impostor (from other subjects) swipes, more accurate verification is achieved than with the other configurations.

Furthermore, an analysis of the iOS configuration shows similar behaviour. Our proposed SwipeFormer with the best feature configuration (touchscreen and background sensors) and Euclidean distance achieves an ERR of 13.70%, while other similarity computation approaches relatively improve this result in terms of EER (Shrunk Covariance: 18.20%, KDE: 38.00%, GMM: 41.00%, OC-SVM: 27.70%, B-SVM: 73.70%). These results demonstrate how classifiers such as SVM are able to separate the different classes (genuine and impostor) with a higher margin, achieving better performance.

In addition, as can be seen in Table 4.3, it is interesting to observe that the iOS configuration reaches, in general, better performance than the Android configuration, achieving in the best case a relative improvement of 45.50% (3.60% EER vs. 6.60% EER). We hypothesise that this improvement achieved on iOS can be produced due to all devices are from the same company (Apple), using similar high-quality accelerometer and gyroscope sensors, contrary to the Android case that contains very different smartphone models in terms of sensors. The quality and calibration of the sensors, and the device's

overall design and hardware integration has been studied in (Franček et al., 2023).

Finally, for completeness, Fig. 4.3 shows the DET curves of the proposed SwipeFormer and the previous touchscreen biometric systems, (Fierrez et al., 2018b) and (Acien et al., 2020), in the two configurations studied (Android and iOS). The best feature configuration is analysed by combining the touchscreen and the background sensors (accelerometer and gyroscope). Overall, we can see a similar behaviour in both configurations, where SwipeFormer with the B-SVM approach outperforms (Fierrez et al., 2018b) with a relative improvement in terms of EER of 158.10% and 89.70% in Android and iOS respectively. A similar trend is observed for (Acien et al., 2020) with an EER relative improvement of 62.70% in Android and 75.50% in iOS. These results show how the correct classifier, such as SVM, helps to better adapt the features extracted by the Transformer to each specific subject. This is consistent with related works that have shown subject-adaptation (Fierrez et al., 2018a) to be very useful in behavioural biometrics (Fierrez et al., 2005).

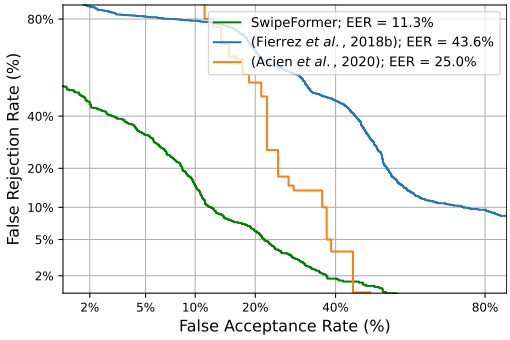
### 4.5.2 Experiment 2: Frank and HuMI Databases

Fig. 4.3 shows the DET curves and the EER results obtained in (Fierrez et al., 2018b) and (Acien et al., 2020) systems, and the proposed SwipeFormer on the public available databases Frank and HuMIDB. All experiments in each database have the same experimental protocols. Furthermore, B-SVM is considered in the similarity computation module of SwipeFormer.

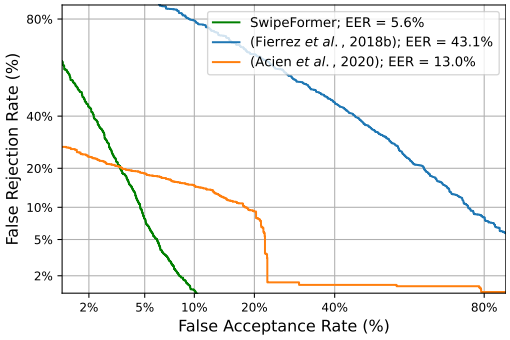
This experiment proves the robustness of the features extracted by our proposed SwipeFormer. Overall, it can be observed how SwipeFormer outperforms previous state-of-the-art approaches in different databases under the same experimental protocol. In particular, for the Frank database (Frank et al., 2012), SwipeFormer achieves an EER

**Figure. 4.3:** DET curves and EER (%) achieved by the proposed SwipeFormer and other state-of-the-art approaches in the literature, i.e., (Fierrez et al., 2018b) and (Acien et al., 2020). The best configuration of the touchscreen and background sensors (accelerometer and gyroscope) is analysed.

**(a)** Frank Database Results




**(b)** HuMI Database Results




of 11.30% in comparison with the 43.60% EER obtained with (Fierrez et al., 2018b) and 25.00% with (Acien et al., 2020) (relative improvements of 74.80% and 56.00% respectively). In addition, for HuMIDB (Acien et al., 2021a), SwipeFormer obtains an EER of 5.60% while (Fierrez et al., 2018b) and (Acien et al., 2020) approaches achieve 43.10% and 13.00% EERs respectively (relative improvements of 88.40% and 61.50%).

The results obtained highlight the significant potential of the proposed Transformer for several reasons. Firstly, the incorporation of GRE allows to introduce in each sample details about its relative position with respect within the sequence, increasing the complexity of the extracted information. Finally, the adoption of a two-stream architecture (Temporal and Frequency Modules) facilitates the extraction of distinct features, obtaining a more complete representation of each sample.

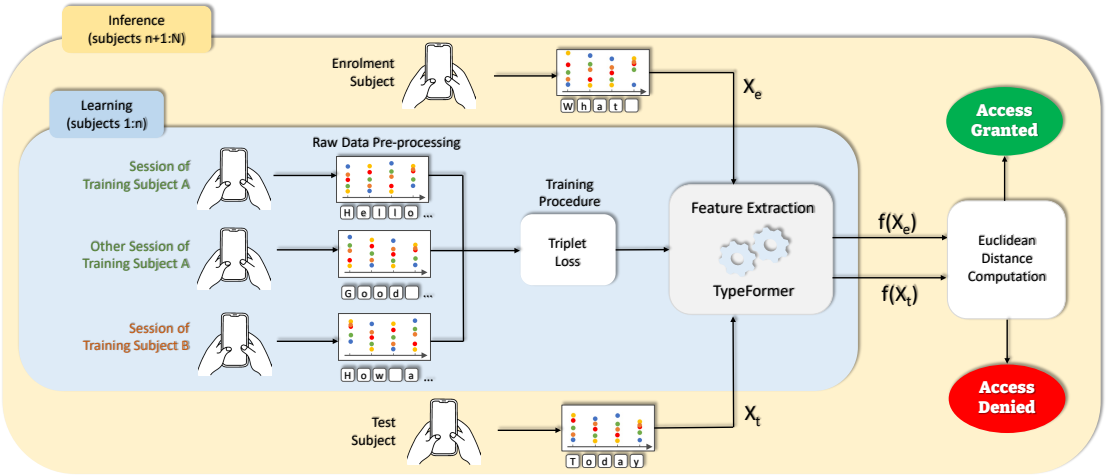
### 4.5.3 Deployment on Real Scenarios

This section analyses the time consumption of the proposed SwipeFormer. It is important to highlight that the training of SwipeFormer, like most DL models, is typically carried out off-line on powerful computers or servers with dedicated GPUs. As indicated

in Fig. 4.1, this corresponds to the training stage. Once the model is trained using large-scale databases, it can be deployed in real-time applications for feature extraction, also known as the inference stage. In addition, it is important to remark that there are various options for storing and running DL models. For instance, the DL model can be stored on a remote server, which receives input data from the mobile device, calculates the score or prediction, and then sends it back to the mobile device to make a decision based on the obtained result. This approach allows the computational burden to be shifted to the server, leveraging its higher processing capabilities, while the mobile device primarily handles input/output communication and decision-making based on the received scores or predictions.

Therefore, in this section we analyse the time consumption of SwipeFormer in the inference stage, simulating the final application scenario. All experiments are carried out using *PyTorch* with an *Intel Core i7-12700K* processor and an *NVIDIA GeForce RTX 3090* graphics card. For reproducibility reasons, we consider in this analysis the publicly available databases Frank DB and HuMIdb. Regarding the Frank DB, SwipeFormer achieves a significantly reduction in time, only 2.11 ms per comparison (on average), surpassing the models presented by (Fierrez et al., 2018b) with 567.88 ms and (Acien et al., 2020) with 7.66 ms. Similarly, in the case of HuMIdb, SwipeFormer outperforms previous state-of-the-art approaches with a time of 0.34 ms per comparison (on average), compared to 3.39 ms and 7.61 ms achieved by (Fierrez et al., 2018b) and (Acien et al., 2020), respectively. These results demonstrate that SwipeFormer not only improves recognition accuracy but also excels in terms of time efficiency compared to previous approaches.

**Figure. 4.4:** Graphical representation of the workflow of TypeFormer, the proposed biometric keystroke free-text verification system.




## 4.6 Application to Other Behavioural Biometrics: Keystroke

In addition to the proposal of Transformer-based systems for gait (Chapter 3) and swipe (Chapter 4) biometrics, we have also conducted an assessment of their applicability to other behavioural biometric traits. Specifically, as part of the PriMa project[1], we collaborated with Giuseppe Stragapede (ESR1 at UAM) to examine the efficacy of Transformers in the field of mobile keystroke dynamics. In this Section, we introduce *TypeFormer*, an innovative Transformer architecture designed for biometric keystroke free-text verification. Additionally, we provide a detailed analysis of the various modules that comprise the final architecture. This Chapter presents a brief overview of the conducted work. For a comprehensive understanding and contextual information, we kindly refer readers to the following articles (Stragapede et al., 2023a,b).

[1]PriMa ITN: `https://www.prima-itn.eu/`.

### 4.6.1 Proposed Approach: TypeFormer

A general description of the proposed keystroke verification system for mobile scenarios is presented in Fig. 4.4. The system operates by first extracting 5 features per character typed in a virtual keyboard on the touchscreen, including hold latency, inter-key latency, press latency, release latency, and the specific key pressed. These features serve as the input for subsequent processing.

The TypeFormer architecture is used for feature extraction and consists of a Temporal Module and a Channel Module, similar to M-GaitFormer (Chapter 3). In both channels, the input sequence is modelled using GRE to preserve the information position.

The Temporal Module encompasses three distinct layers sets, each comprising different numbers of layers: $N$, $R$, and $M$. The $N$ and $M$ layers share identical structure, comprising a multi-head Self-Attention mechanism and a multi-scale keystroke LSTM-RNN layer. Additionally, similar to M-GaitFormer, $R$ recurrent layers are incorporated between the $N$ and $M$ layers.

The Channel Module, on the other hand, takes the transposed input sequence and applies the GRE modelling technique. It includes $H$ layers, analogous to the $N$ and $M$ layers of the Temporal Module.

### 4.6.2 Databases Description and Experimental Setup

For the purpose of development and evaluation, the Aalto mobile keystroke database (Palin et al., 2019) was chosen as it is a very popular and large-scale database. Additionally, in order to conduct a cross-database evaluation, the Aalto desktop keystroke

**Table 4.4:** Intra-database evaluation: System performance results in terms of EER for the final evaluation dataset of the Aalto mobile database (Palin et al., 2019).

| **Sequence Length $L$** | **Model** | **Number of Enrolment Sessions $E$** | | | | |
|---|---|---|---|---|---|---|
| | | **1** | **2** | **5** | **7** | **10** |
| **30** | (Acien et al., 2021b) | 14.20 | 12.50 | 11.30 | 10.90 | 10.50 |
| | **TypeFormer** | **9.48** | **7.48** | **5.78** | **5.40** | **4.94** |
| **50** | (Acien et al., 2021b) | 12.60 | 10.70 | 9.20 | 8.50 | 8.00 |
| | (Stragapede et al., 2023a) | 6.99 | - | 3.84 | - | 3.15 |
| | **TypeFormer** | **6.17** | **4.57** | **3.25** | **2.86** | **2.54** |
| **70** | (Acien et al., 2021b) | 11.30 | 9.50 | 7.80 | 7.20 | 6.80 |
| | **TypeFormer** | **6.44** | **5.08** | **3.72** | **3.30** | **2.96** |
| **100** | (Acien et al., 2021b) | 10.70 | 8.90 | 7.30 | 6.60 | 6.30 |
| | **TypeFormer** | **8.00** | **6.29** | **4.79** | **4.40** | **3.90** |

database (Dhakal et al., 2018), Clarkson II database (Murphy et al., 2017), and Buffalo database (Sun et al., 2016) were also utilised.

### 4.6.3 Experimental Results

Table 4.4 presents the intra-database verification results obtained by TypeFormer for different sequence lengths, denoted as $L$, and different number of enrolment sessions, denoted as $E$. To provide a comprehensive comparison with recent state-of-the-art systems, we include the results achieved by TypeNet as reported in (Acien et al., 2021b), as well as our preliminary study conducted on the same database (Aalto mobile database) (Stragapede et al., 2023a).

The verification results demonstrate that TypeFormer consistently outperforms previous approaches across all configurations. Specifically, the average relative improvement of TypeFormer, considering all cases in the table ($E = 1, 2, 5, 7, 10$ and $L = 30, 50, 70, 100$), is 47.30% in comparison to TypeNet (Acien et al., 2021b), an LSTM RNN-based system.

Furthermore, focusing solely on TypeFormer, it is observed that the EER values

**Table 4.5:** Cross-Database Evaluation: EER (%) achieved by TypeFormer in comparison with TypeNet (Acien et al., 2021b). The databases considered are Aalto Mobile (development set) (Palin et al., 2019), Aalto Desktop (Dhakal et al., 2018), Clarkson II (Murphy et al., 2017), and SUNY Buffalo (free-text and transcripted text) (Sun et al., 2016) (all in the desktop scenario). *Experiments using all the available data per subject.

| **Evaluation Database** | **(Acien et al., 2021b)** | **TypeFormer** |
|---|---|---|
| Aalto Mobile | 9.20 | 3.25 |
| Aalto Desktop | 21.40 | 15.02 |
| Clarkson II | 36.60 | 27.83 |
| Clarkson II* | 33.00 | 25.34 |
| SUNY Buffalo (Free) | 33.20 | 22.39 |
| SUNY Buffalo (Transcript) | 32.80 | 23.40 |

consistently decrease as the number of enrolment sessions $E$ increases. This trend remains consistent across different sequence lengths $L$. Notably, the rate of improvement is more pronounced when transitioning from $E = 1$ to $E = 5$ sessions, resulting in an approximately 50.00% relative improvement.

Finally, for completeness, Table 4.5 shows the cross-database results obtained by TypeFormer ($E = 5$, $L = 50$) on various databases and keystroke scenarios that were not included during the development phase (Aalto mobile keystroke database). This experimental analysis serves to evaluate the generalisation ability and robustness of the features extracted by TypeFormer.

By examining the results, it becomes apparent that there is a notable degradation in performance when considering different databases. This underscores the significance of this aspect in real-life applications, and it should not be underestimated. It is crucial to emphasise that we have not employed any fine-tuning strategies for the model. Nonetheless, the proposed TypeFormer exhibits significant mitigation of such performance degradation when compared to (Acien et al., 2021b), yielding an average absolute improvement of 8.60% EER across the cross-database evaluation cases under consideration.

## 4.7 Conclusions

The present Chapter has introduced SwipeFormer, a novel touchscreen verification system based on Transformers. To the best of our knowledge, this is the first attempt to apply Transformers to touchscreen biometrics.

SwipeFormer consists of two modules: *i)* a feature extractor based on a Transformer architecture, trained with the development data acquired from the touchscreen and the background sensors of the mobile device in the learning stage; and *ii)* a similarity computation module (Euclidean distance, Shrunk Covariance, KDE, GMM, OC-SVM, and B-SVM), which provides a final verification with a evaluation dataset (inference stage). Similar to M-GaitFormer, described in previous Chapter, SwipeFormer contains two modules (Temporal and Frequency) with a GRE, multi-head Self-Attention mechanism, and CNNs.

Two experiments are carried out considering an in-house database collected under unconstrained conditions and two of the most popular public databases in touchscreen biometric verification collected under constrained conditions (Frank DB and HuMIDB). For the publicly available databases, the same experimental protocol proposed in the literature was considered. Regarding the experimental results, SwipeFormer outperforms state-of-the-art systems in all databases, achieving EER of 3.60%, 11.30%, and 5.60% in our in-house database, Frank DB, and HuMIDB respectively.

In addition, Transformers have also been adapted to mobile keystroke dynamic verification, presenting TypeFormer. TypeFormer has also outperformed previous approaches presented in the literature on popular databases, reducing also the traditional performance gap existing between mobile free-text and desktop fixed-text scenarios. Finally, we also analyse the behaviour of the model with different experimental configurations

such as the length of the keystroke sequences and the amount of enrolment sessions.

Finally, we conclude this Chapter by demonstrating the effectiveness of Transformers in various biometric scenarios (swipe and keystroke dynamics). It is important to emphasise that considering the unique characteristics of each trait, it becomes crucial to adapt these architectures accordingly for each specific task. We have successfully presented significant breakthroughs that have managed to improve the state of the art.

# Part III

# Mobile Biometric Privacy

# Chapter 5

# GaitPrivacyON: Privacy-preserving Mobile Gait Biometrics

## 5.1 Introduction

### 5.1.1 Generic Privacy Preservation

In the previous research Chapters 3 and 4, we have explored the significance of ensuring security for mobile devices, emphasising the use of behavioural biometrics. However, despite the progress we have made, a general concern that continues to demand attention is the privacy of data in these devices (Melzi et al., 2022a).

Although mobile behavioural biometrics has gained popularity, data collected through mobile devices may contain personal and sensitive information, including demographic

data (such as gender, age, and ethnicity) or the activities performed by individuals, among others (Tolosana et al., 2022a). Consequently, concerns regarding the invasion of personal privacy have arisen in relation to this technology.

In response to privacy concerns, our thesis presents an innovative approach to safeguard the personal data of individuals when employing biometrics on mobile devices. Specifically, we delve into the realm of gait privacy. The rationale behind this focus is grounded in the recognition task exploration within this biometric trait. By extending our investigation to encompass privacy considerations in a domain we have already studied, we aim to provide a comprehensive demonstration of privacy measures in tandem with the previously examined recognition aspects.

### 5.1.2 Privacy in Gait Recogniton

Privacy concerns in gait recognition stem from the unique and identifiable nature of gait patterns, posing risks such as identity theft and unauthorised tracking. Deploying gait recognition in public spaces raises ethical issues, as individuals may be monitored without their consent, and the technology may inadvertently disclose sensitive attributes like health conditions. Additionally, the potential for biometric spoofing and the need for robust data security measures underscore the importance of balancing the benefits of gait recognition with safeguarding individual privacy. To address these challenges, it is essential for researchers and practitioners to implement privacy-preserving techniques, including encryption and anonymisation, and to ensure compliance with privacy regulations. Achieving this balance is crucial for responsible development and deployment of gait recognition technologies.

In particular, in order to overcome these limitations, in this chapter GaitPrivacyON is presented. By employing unsupervised learning techniques, this approach offers an

effective means of preserving privacy in the context of mobile behavioural biometrics.

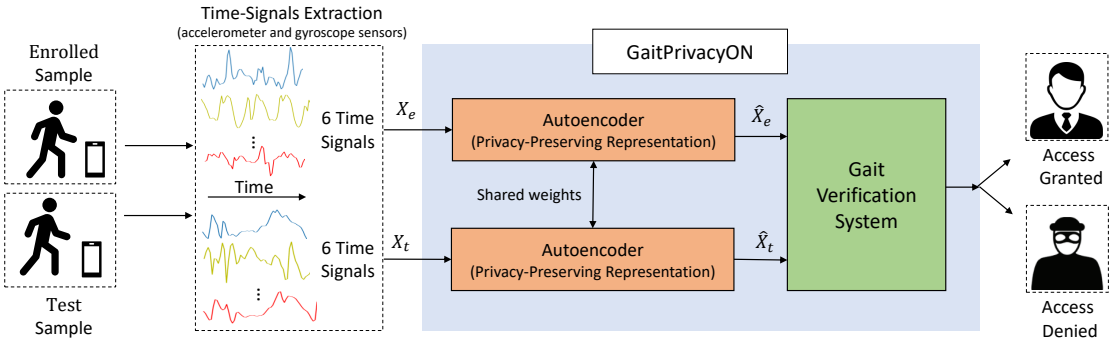
This Chapter is structured as follows: Section 5.2 describes the architecture of GaitPrivacyON, including the different Autoencoders architectures, the gait verification system, and training. Section 5.3 describes the databases while Section 5.4 provides a description of the system details and experimental protocol. Section 5.5 describes the results achieved. Finally, Section 5.6 draws the final conclusions. This Chapter is based on the following publication: (Delgado-Santos et al., 2022b).

## 5.2 Proposed Approach: GaitPrivacyON

Fig. 5.1 provides a general representation of GaitPrivacyON, our proposed privacy-preserving approach for mobile behavioural biometrics. GaitPrivacyON considers a Siamese architecture that is used to learn the similarity between two different biometric templates from the same subject (genuine) or from different subjects (impostor) (De Luisa et al., 2019). First, we consider as input the accelerometer and gyroscope time sequences captured by the mobile device. After that, GaitPrivacyON comprises two modules: *i)* two convolutional Autoencoders with shared weights that transform the raw biometric data into a new privacy-preserving representation (Section 5.2.1); and *ii)* a mobile gait verification system based on the combination of CNNs and RNNs with a Siamese architecture (Section 5.2.2). For the training, we adapted the key aspects presented in the image style transformation field (Zhang et al., 2021). The details are explained in Section 5.2.3. GaitPrivacyON is an improved adaptation of the approach presented in (Zhang et al., 2021). We clarify next the main changes:

- GaitPrivacyON is based on gait biometric verification while the approach presented in (Zhang et al., 2021) is based on activity recognition. As a result, our approach

**Figure. 5.1:** Diagram of GaitPrivacyON, which comprises two modules: *i)* two Autoencoders that are in charge of removing automatically the sensitive data; and *ii)* a gait verification system. Time signals extracted from the accelerometer and gyroscope sensors of the mobile devices are considered as input to GaitPrivacyON.




$X_e$: Enrolled sample, $X_t$: Test sample, $\widehat{X}_e$: Transformed enrolled sample, $\widehat{X}_t$: Transformed test sample.

focuses on verification (1:1) rather than identification (1:N).

- Regarding the Autoencoders considered in GaitPrivacyON (see details in Section 5.2.1), while TransNet only has a single Autoencoder, two Autoencoders are considered in GaitPrivacyON, sharing their weights through a Siamese architecture. In addition, in order to extract more discriminative features and improve the training, we have considered batch normalisation and increased the complexity of the network using more convolutional layers.

- The Gait Verification System proposed in GaitPrivacyON (see details in Section 5.2.2) has several differences compared with LossNet (Zhang et al., 2021). LossNet is based only on convolutional layers. The system presented in this work considers both convolutional and recurrent layers following the state of the art in gait biometrics (Zou et al., 2020). This improves the performance of the system and makes our system more robust.

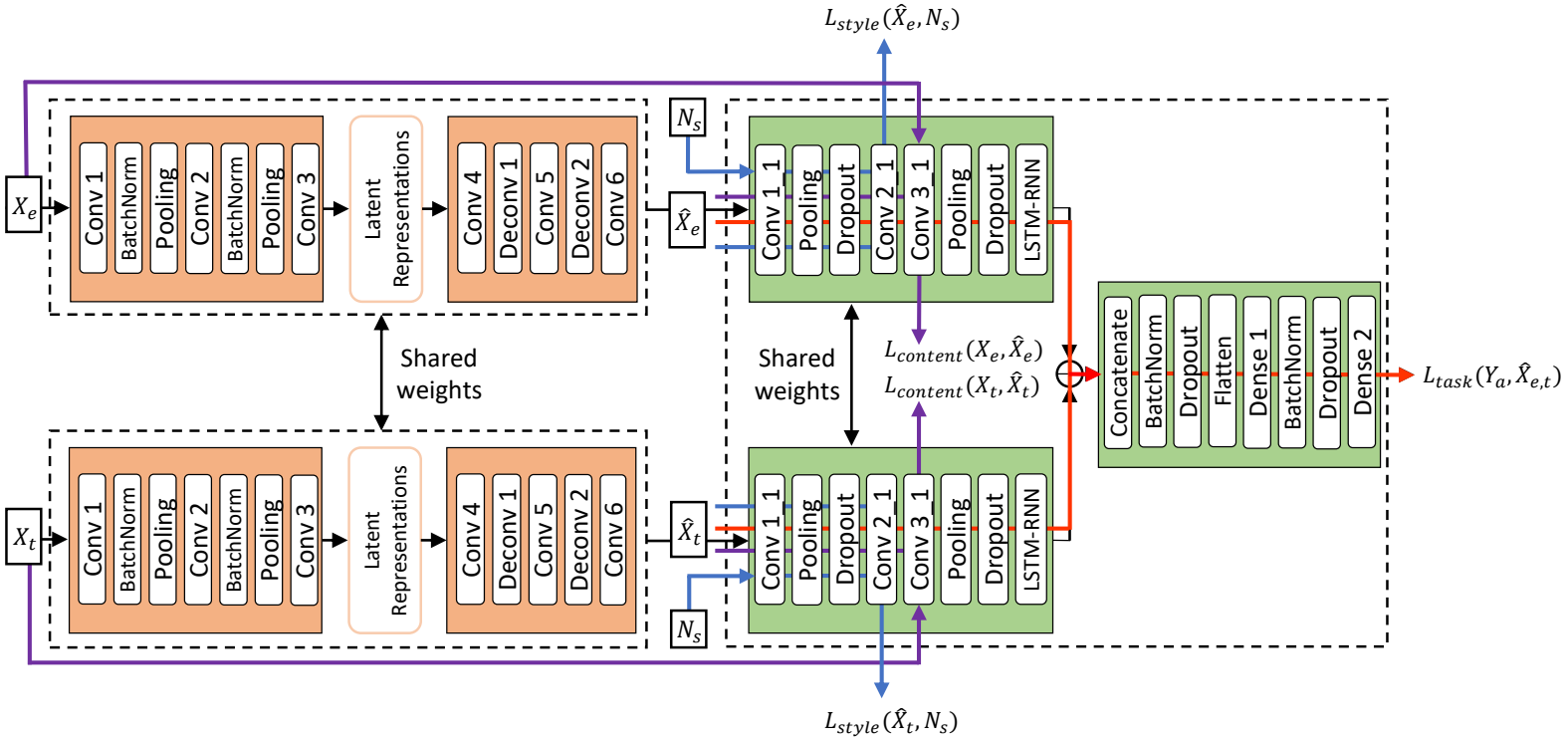
### 5.2.1 Autoencoders

Fig. 5.2 (orange colour) provides a graphical representation of the proposed module. It comprises two convolutional Autoencoders with the same architecture and shared weights. The inputs for each Autoencoder are an enrolled sample ($X_e$) and a test sample ($X_t$), while the outputs are the transformed versions of the enrolled sample ($\widehat{X_e}$) and the test sample ($\widehat{X_t}$), respectively. The architecture of both is composed of a sequence of $1 \times 3$ convolutional filters, coupled with ReLU activation functions. In the encoder, after each convolutional layer, batch normalisation and $1 \times 2$ max-pooling layers are used to decrease the size of the activation map. In the decoder, after each convolutional layer, a deconvolutional layer is used with $1 \times 3$ strides of the convolution. The activation function of the last convolutional layer is linear. In GaitPrivacyON, the training loss function of the main task ($\mathcal{L}_{task}$) is in charge of training the Autoencoders to extract useful transformed data ($\widehat{X}$). This training loss function is considered together with the loss of content ($\mathcal{L}_{content}$), responsible for retaining recognition information, and the training loss function of style ($\mathcal{L}_{style}$), which removes sensitive data by introducing uniform random noise ($N_s$).

### 5.2.2 Gait Verification System

Fig. 5.2 (green colour) provides a graphical representation of the architecture proposed for gait verification ($\varphi$). In particular, we have adapted the approach originally presented in (Zou et al., 2020) to our specific case (privacy-preserving gait verification). It is based on a novel Siamese architecture with two inputs: transformed enrolled sample ($\widehat{X_e}$) and transformed test sample ($\widehat{X_t}$). The inputs are reshaped including one new dimension. Unlike the method proposed in (Zou et al., 2020), the architecture is composed of a sequence of $1 \times 3$ two-dimensional convolutional filters, coupled with ReLU activation

**Figure. 5.2:** Architecture and training losses ($\mathcal{L}_{content}$, $\mathcal{L}_{style}$, $\mathcal{L}_{task}$) considered in GaitPrivacyON




$X_e, X_t$: Raw time signals; $\widehat{X_e}, \widehat{X_t}$: Transformed time signals; $N_s$: Random noise; $Y_a$: label of the gait verification task.

functions. After 3 convolutional layers, batch normalisation, $1 \times 2$ max-pooling, and dropout with a probability of 0.5 are used. A reshaping layer is included to return to the shape of the time domain signals following by a bi-directional LSTM layer with 50 units. The dense layer has a size of 400 with a sigmoid activation function.

### 5.2.3 Training

GaitPrivacyON is trained following the idea proposed in the image style transformation field (Johnson et al., 2016). One image can be divided into two parts: *i)* the *content*, i.e., what is in the image, and *ii)* the *style*, i.e., how the image is illustrated. In our particular application of gait biometric verification, the content is the unique information that allows to verify the identity of the subject whereas the style is the sensitive information of the subject that can be considered for other purposes not related to the recognition. This sensitive information may include the person's gender, age, ethnicity, or the activity the subject is performing while using mobile devices (Iwasawa et al., 2017).

Following this idea, three different loss functions have been considered from the work presented in (Zhang et al., 2021): *task loss* ($\mathcal{L}_{task}$), *content loss* ($\mathcal{L}_{content}$), and *style loss* ($\mathcal{L}_{style}$).

The *task loss* ($\mathcal{L}_{task}$) helps the system to maintain its usefulness in the main task of gait verification. We consider a categorical cross-entropy that compares the transformed data ($\widehat{X}$) with the raw biometric data ($X$). The *task loss* can be defined as:

$$\mathcal{L}_{task}(Y_a, \widehat{X}) = -Y_a log(\varphi(\widehat{X})) \tag{5.1}$$

where $Y_a$ and $\varphi(\widehat{X})$ are the label and the predicted probability of the gait verification task respectively.

The *content loss* ($\mathcal{L}_{content}$) measures the content (i.e., the recognition information) that the transformed data ($\widehat{X}$) and the raw biometric data ($X$) have in common. For this aim, we use the Euclidean distance to compare the feature maps provided by the $i$-layer of the $\varphi$ network when using both the raw biometric data and the transformed data as input. In our case, we use the feature maps obtained behind the *Conv3_1* layer in Fig. 5.2. This was decided experimentally. The *content loss* is defined as:

$$\mathcal{L}^{i}_{content}(X, \widehat{X}) = \frac{1}{C_i H_i W_i} \left\| \varphi_i(\widehat{X}) - \varphi_i(X) \right\|_2^2 \tag{5.2}$$

where $i$ is the layer and $C_i \times H_i \times W_i$ is the shape of the feature map obtained after this layer. Comparing feature maps ensures that the content of the raw biometric data and the transformed data are similar but do not have to be identical.

The *style loss* ($\mathcal{L}_{style}$) is responsible for maintaining the transformed data ($\widehat{X}$) unstyled, thus avoiding the extraction of any sensitive information automatically. For this purpose, we want to modify the style of the data by uniform random noise ($N_s$) with range [-20, 20] as done by (Zhang et al., 2021). We consider the Gram matrix ($G$) to measure the style differences between feature representations. Random noise is introduced as the new domain, avoiding using any information from the sensitive data for its protection, creating an unsupervised learning framework. For this aim, both the transformed data and the random noise are fed into the trained gait verification system with the weights frozen. After that, the Gram Matrices of the feature maps obtained as output of the $i$-layer are compared. The Gram Matrix can be defined as:

$$G_i(X)_{c,c'} = \frac{1}{C_i H_i W_i} \sum_{h=1}^{H_i} \sum_{w=1}^{W_i} \varphi_i(X)_{h,w,c} \varphi_i(X)_{h,w,c'} \tag{5.3}$$

where the shape of $\varphi_i(X)$ is $C_i \times H_i \times W_i$ and the shape of its Gran matrix ($G_i^{\varphi}$) is

$|C_i| \times |C_i|$. $\varphi_i(X)$ can be interpreted as $C_i$ dimensional features for each $H_i \times W_i$ point, where $c$ and $c'$ are two different dimensions.

The *style loss* measures the dissimilarity in style using the Frobenius squared norm of the difference of the Gram matrices of the transformed data ($\widehat{X}$) and the random noise ($N_s$). In our case, we have decided to use the feature maps obtained behind *Conv2_1* in Fig. 5.2. The *style loss* can be defined as:

$$\mathcal{L}^{i}_{style}(\widehat{X}, N_s) = \left\| G^{\varphi}_{i}(\widehat{X}) - G^{\varphi}_{i}(N_s) \right\|^2_F \tag{5.4}$$

where $F$ denotes the Frobenius squared norm. By using deeper layers, the extracted features will be more similar.

The final loss function of GaitPrivacyON ($\mathcal{L}_{total}$) would be a weighted sum of the losses $\mathcal{L}_{task}$, $\mathcal{L}_{content}$, and $\mathcal{L}_{style}$:

$$\mathcal{L}_{total} = \alpha\mathcal{L}_{task} + \beta\mathcal{L}_{content} + \gamma\mathcal{L}_{style} \tag{5.5}$$

where $\alpha + \beta + \gamma = 1$.

## 5.3 Databases Description

Three popular public databases used for research in privacy-preserving gait recognition on mobile devices are considered in the evaluation framework of the present study: *i)* MotionSense (Malekzadeh et al., 2018), *ii)* MobiAct (Vavoulas et al., 2016), and *iii)* OU-ISIR (Ngo et al., 2014). In particular, OU-ISIR database has been already introduced in Chapter 3. However, MotionSense and MobiAct databases are explained in this Chapter.

### 5.3.1 MotionSense Database

The MotionSense database (Malekzadeh et al., 2018) comprises accelerometer and gyroscope data collected with an iPhone 6s. Data from a total of 24 subjects, with information on gender, age, height, and weight, were obtained. The data was acquired while the subjects performed 4 different activities (walking up and down stairs, jogging, and walking). All the subjects had the mobile phone fixed in the front pocket of the trousers.

### 5.3.2 MobiAct Database

The MobiAct database (Vavoulas et al., 2016) comprises accelerometer, gyroscope and magnetometer data collected using a Samsung Galaxy S3. A total of 56 subjects performing the same 4 activities (walking up and down stairs, jogging, and walking) were captured. Data on gender, age, height, and weight of the subjects were acquired. Unlike the previous database, subjects had a free choice of placement of their device, simulating a realistic scenario.

## 5.4 Experimental Setup

GaitPrivacyON considers two main tasks: *i)* gait biometric verification, and *ii)* privacy-preserving information. First, GaitPrivacyON is trained in order to ensure these two tasks at the same time. In addition, to check the sensitive information contained in the biometric data, auxiliary ML systems are implemented. In our scenario, this includes detecting the gender and activity of an individual while using the mobile device (i.e., gender and activity inference systems).

**Table 5.1:** Architecture of the gender and activity inference systems. Prob- Probability. m- number of signals. SAC- Sensitive Attribute Classes.

| **Layer** | **Input Size** $(H \times W \times F)$ | **Kernel** $(H \times W)$ | **Padding** | **Activation** | **Prob** |
|---|---|---|---|---|---|
| Conv1_1 | m×100×1 | 1×3 | Valid | Relu | - |
| Conv1_2 | m×98×16 | 1×3 | Valid | Relu | - |
| Batch_1 | m×96×16 | - | - | - | - |
| Pool_1 | m×96×16 | 1×2 | Valid | - | - |
| Drop_1 | m×48×16 | - | - | - | 0.5 |
| Conv2_1 | m×48×16 | 1×5 | Valid | Relu | - |
| Batch_2 | m×44×32 | - | - | - | - |
| Pool_2 | m×22×32 | 1×2 | Valid | - | - |
| Drop_2 | m×22×32 | - | - | - | 0.5 |
| Dense_1 | m×100 | - | - | - | - |
| Batch_3 | m×100 | - | - | - | - |
| Drop_3 | m×100 | - | - | - | 0.5 |
| Dense_2 | m×SAC | - | - | - | - |

### 5.4.1 GaitPrivacyON System Details

Regarding the training procedure, GaitPrivacyON first trains only the gait verification system using the raw biometric data ($X$) from the development dataset. For this first stage, binary cross-entropy is considered for the loss function. After that, we train our proposed GaitPrivacyON approach (only the Autoencoders module, the weights of the gait verification system are frozen) using the same development dataset. In this second stage, the total loss function ($L_{total}$) considered in GaitPrivacyON is a weighted sum of the losses $L_{task}$, $L_{content}$, and $L_{style}$, as described in Section 5.2. The specific details of the system details are provided in Section 5.4.2 and of the development and final evaluation datasets are provided in Section 5.4.3.1 and Section 5.4.3.2.

### 5.4.2 Gender and Activity Inference Systems Details

Table 5.1 shows the architecture of the proposed gender and activity inference systems. The input data is in the same shape as in GaitPrivacyON. The architecture is composed

of a sequence of $1 \times 3$ convolutional filters, coupled with ReLU activation functions. After a series convolutional layers, batch normalisation, $1 \times 2$ max-pooling, and dropout with a probability of 0.5 are used. The dense layer has a size of 100. For the gender recognition system, a sigmoid activation function is considered whereas softmax is considered for the activity recognition system. Finally, cross-entropy is used for the loss function.

### 5.4.3 Experimental Protocol

#### 5.4.3.1 MotionSense & MobiAct Databases

Our approach is trained with accelerometer and gyroscope time signals combined from both MotionSense and MobiAct databases. A total of 80 subjects (i.e., 24 from MotionSense and 56 from MobiAct) performing 4 different activities (walking up and down stairs, jogging, and walking) are considered in the experimental framework. The total database consists of 55 male and 25 female subjects. In both databases the sample frequency has been normalised to a mean of 0 and a standard deviation of 1, with a sampling frequency of 50 Hz. Each time signal comprises 100 samples. Also, we consider time windows of 2 seconds with an overlapping ratio of 75%. The total database is divided into development and evaluation datasets, which contain different subjects with random selection. The development dataset, used for the training of GaitPrivacyON, has 70 subjects (85% of the subjects have been used for training and the remaining part for validation). After training, the remaining 10 unseen subjects are used for the final evaluation. Regarding the gender and activity inference systems (see Section 5.4.2), we consider the same development and evaluation datasets described above, balancing the number of male and female subjects to avoid bias (5 males and 5 females in the final evaluation set). All subject data contain the same 4 activities.

#### 5.4.3.2 OU-ISIR Database

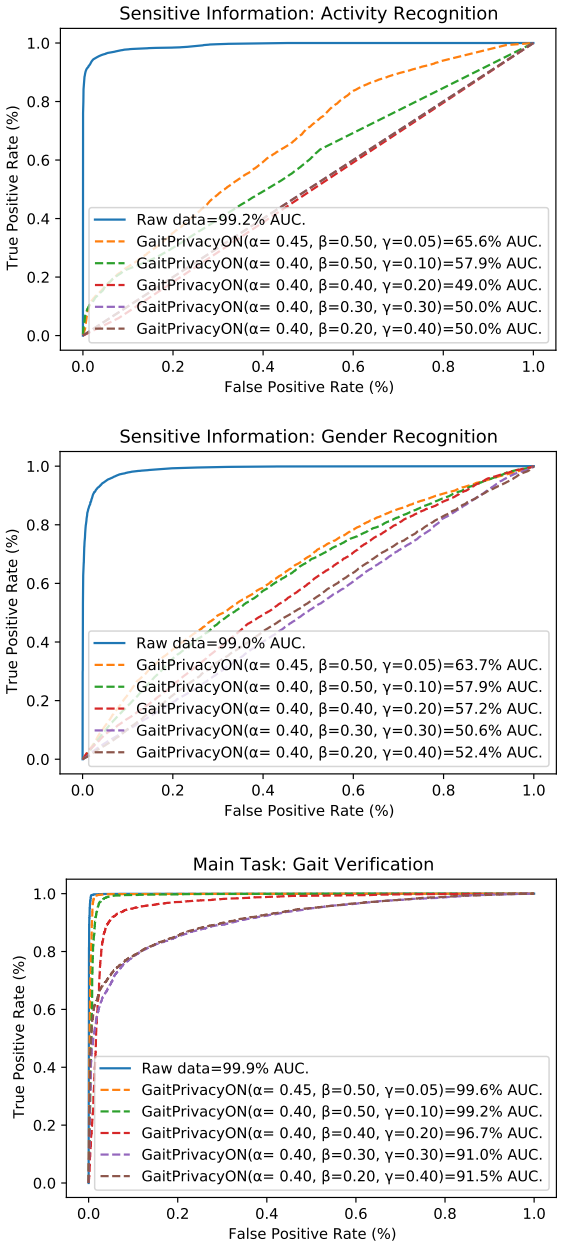
GaitPrivacyON is trained with accelerometer and gyroscope time signals using the right-position inertial measurement unit, as it is more reliable according to (Ngo et al., 2014). In the scenario of performing 4 different activities (two flat walking, slope-up walking, and slope-down walking), there are 492 subjects available (256 males and 236 females). The data have been normalised with a mean of 0 and a standard deviation of 1, with a sampling frequency of 100 Hz. Each time signal has a time window of 1 second, which is defined as 100 samples, and an overlapping between time windows of 75%. This database is divided into development and evaluation, which comprises different subjects with random selection. For the training of GaitPrivacyON, the development dataset contains 80% of the subjects (312 for training and 80 for validation). After the training, the remaining 20% of the subjects (100 unseen subjects) are used for the final evaluation. Regarding the gender and activity inference systems (described in Section 5.4.2), we consider the same development and evaluation datasets described above, balancing the number of male and female subjects to avoid bias (50 males and 50 females in the final evaluation set). All subjects contain the same 4 activities.

## 5.5 Experimental Results

### 5.5.1 Gender and Activity Inference from Raw Biometric Data

In this first experiment we analyse the ability of ML systems to infer sensitive information of the subject from the raw biometric data.

**Figure. 5.3:** ROC curves and AUC (%) results on the MotionSense and MobiAct evaluation dataset for the two scenarios considered: *i)* Raw biometric data ($X$), and *ii)* GaitPrivacyON ($\widehat{X}$). Different parameters $(\alpha, \beta, \gamma)$ of GaitPrivacyON are tested in order to evaluate the results of the main task (gait verification) and the privacy-preserving information of the subject (activity, the mean of the AUC for each activity, and gender recognition). Activity recognition system (solid curve) and GaitPrivacyON (dashes curves).



#### 5.5.1.1 MotionSense & MobiAct Databases

Fig. 5.3 (top) shows the Receiver Operating Characteristic (ROC) curve together with the AUC of the activity recognition system (solid curve). The proposed system achieves 99.20% AUC, differentiating the activity (walking up and down stairs, jogging, and walking) with precision. For activity recognition, a dedicated analysis has been conducted for each of the four activities. The AUC has been thoroughly examined for each activity, and an average has been computed based on the results of these individual assessments.

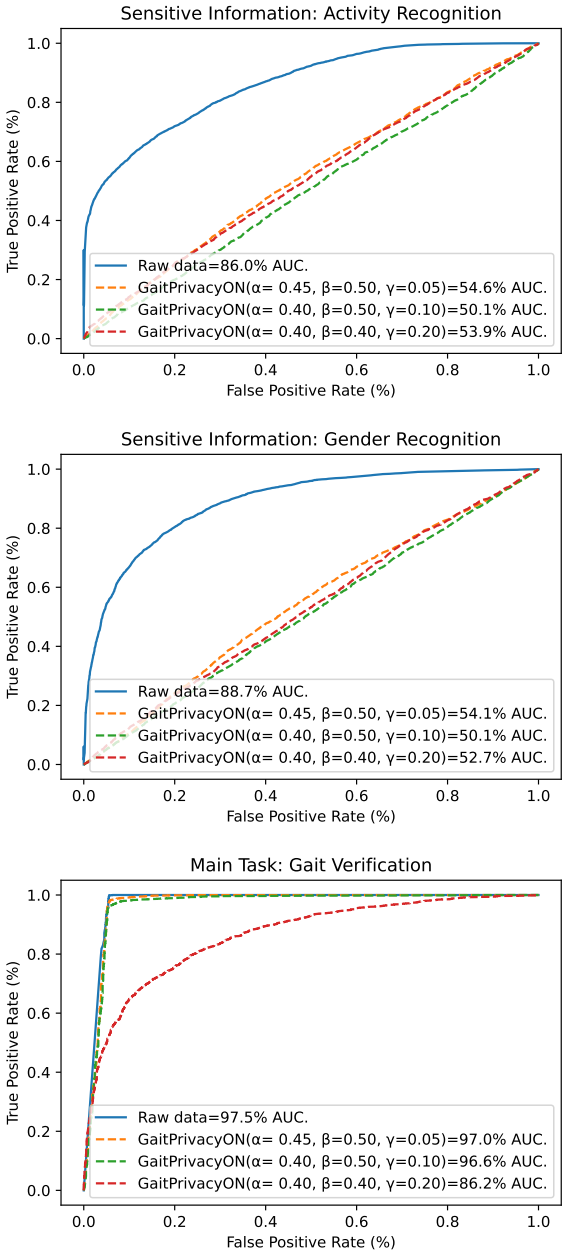
Second, we analyse the results achieved by the proposed gender recognition system. The system has two clases: male and female. Fig. 5.3 (middle) shows the ROC curve together with the AUC achieved by the gender recognition system (solid curve). As in the case of the activity task, the gender recognition system is able to differentiate the gender with 99.00% AUC.

#### 5.5.1.2 OU-ISIR Database

Fig. 5.4 (top) shows the ROC curve together with the AUC result achieved by the activity recognition system (solid curve). For activity recognition, a dedicated analysis has been conducted for each of the four activities. Similar to the MotionSense and MobiAct databases, the system is able to achieve accurate results with 86.00% AUC. Regarding the gender recognition, see Fig. 5.4 (middle), good results are also achieved with 88.70% AUC.

These results support the ability of ML systems to infer sensitive information of the subjects from the raw biometric data ($X$), which might be considered as an invasion of personal privacy. The following experiments analyse the results achieved by the proposed GaitPrivacyON approach considering the privacy-preserving domain ($\widehat{X}$).

**Figure. 5.4:** ROC curves and AUC (%) results on the OU-ISIR evaluation dataset for the two scenarios considered: *i)* Raw biometric data ($X$), and *ii)* GaitPrivacyON ($\widehat{X}$). Different parameters $(\alpha, \beta, \gamma)$ of GaitPrivacyON are tested in order to evaluate the results of the main task (gait verification) and the privacy-preserving information of the subject (activity, the mean of the AUC for each activity, and gender recognition). Activity recognition system (solid curve) and GaitPrivacyON (dashes curves).



### 5.5.2 GaitPrivacyON

Three different parameters can be configured in the training process of GaitPrivacyON to control the data transformation and the trade-off between the utility of the gait verification (main task) system and the sensitive information of the subject (activity and gender): $\alpha$ (*task loss* parameter), $\beta$ (*content loss* parameter), and $\gamma$ (*style loss* parameter).

#### 5.5.2.1 MotionSense & MobiAct Databases

We first analyse the results achieved in the main task, mobile gait verification. Fig. 5.3 (bottom) shows the ROC curves together with the AUC results of the gait biometric verification system. Analysing the traditional approach (i.e., using the raw biometric data ($X$)) the gait verification system is able to achieve accurate results with 99.90% AUC over the final evaluation dataset. However, from this traditional approach it is also possible to extract sensitive subject information, 99.20% AUC for activity recognition and 99.00% AUC for gender recognition.

The results achieved by GaitPrivacyON in the main task (gait verification) can be seen in Fig. 5.3 (bottom). In general, we can see different AUC results depending on the values of the training parameters (including symbols), ranging from 99.60% AUC to 91.00% AUC. The selection of these parameters affects in the performance of the activity and gender extraction.

Fig. 5.3 (top) shows the ROC curves together with the AUC results achieved by GaitPrivacyON in the activity recognition task (dashes curves) when $X$ is replaced by $\widehat{X}$. It can be observe that the AUC results decrease as $\gamma$ increases, achieving a result close to random (49.02% AUC) when $\gamma = 0.20$.

A similar trend is also observed for the gender recognition task. Fig. 5.3 (middle) shows the ROC curves together with the AUC results achieved by GaitPrivacyON in the gender recognition task (dashes curves). Again, it can be observe that the AUC results decrease as $\gamma$ increases, achieving a result close to random (50.60% AUC) when $\gamma = 0.30$.

As a result, when the transformed data ($\widehat{X}$) provided by GaitPrivacyON achieves AUC values close to random (50.00%) in the sensitive subject information tasks, it will be assumed that privacy-preserving results are achieved, while the AUC of the gait verification task hardly decreases. Therefore, we select as the optimal configuration parameters $\alpha = 0.40$, $\beta = 0.40$, $\gamma = 0.20$, as the results in the gait biometric verification task barely decrease (3.15% AUC), while results close to random are achieved for both activity (49.00% AUC) and gender (57.20% AUC).

#### 5.5.2.2 OU-ISIR Database

Fig. 5.4 (bottom) shows the ROC curves together with the AUC results of the gait biometric verification system. Using the raw biometric data ($X$) of the final evaluation dataset, the gait verification system is able to achieve accurate results with 97.50% AUC. As in the previous case, with this traditional approach it is possible to extract sensitive information such as activity (86.00% AUC) and gender (88.70% AUC). For the OU-ISIR database, the best parameter configuration of GaitPrivacyON is $\alpha = 0.40$, $\beta = 0.50$, $\gamma = 0.10$. In this case, GaitPrivacyON achieves AUC results close to 50.00% for both activity and gender recognition, while keeping a similar performance on gait verification compared with the traditional approach (97.50% AUC vs. 96.60% AUC).

### 5.5.3 Comparison with the State of the Art

Analysing MotionSense and MobiAct databases together, GaitPrivacyON is able to decrease the AUC in the gender task (sensitive information) from 99.00% to 57.20% whilest only reducing the performance from 99.90% AUC to 96.70% AUC in gait verification (main task). Moreover, using the OU-ISIR database, GaitPrivacyON also achieves robust results, decreasing the AUC from 88.70% to 50.10% in gender recognition while keeping similar AUC results in the main task, from 97.50% to 96.60%. In comparison to our work, the approach presented in (Garofalo et al., 2019) using the OU-ISIR database decreased the F1-score in the gender recognition task from 73.00% to 52.00% while worsening the accuracy from 90.90% to 85.30% in the gait verification task. However, it is important to note that their method considers supervised learning, while GaitPrivacyON is based on unsupervised learning.

Finally, for completeness, we highlight other approaches focused on the privacy-preserving of time sequences (Boutet et al., 2021; Hajihassnai et al., 2021; Zhang et al., 2021), although the main topic of these studies is different, i.e., activity recognition. A similar trend can be observed when protecting sensitive information such as the age and identity of the person.

## 5.6 Conclusions

This Chapter has presented GaitPrivacyON, a novel mobile gait biometric verification approach that provides accurate recognition results while preserving the privacy of the subject. One of the main advantages of the approach is that the first module (convolutional Autoencoders) is trained in an unsupervised way, without specifying the sensitive attributes of the subject to protect. We have performed an in-depth quantitative analysis

of GaitPrivacyON over three popular databases in the field of gait recognition, MotionSense (Malekzadeh et al., 2018), MobiAct (Vavoulas et al., 2016), and OU-ISIR (Ngo et al., 2014). Our model is able to obtain good results, as the gait biometric verification task barely decreases (3.20% AUC with MotionSense and MobiAct databases and 0.90% with OU-ISIR database) while results close to random are achieved in both activity and gender (∼50.00% AUC) tasks. In conclusion, GaitPrivacyON increases the protection of the sensitive data (e.g., activity and gender) with unsupervised learning whilest being able to maintain the accuracy of the gait biometric verification task.

# Part IV

# Conclusions and Future Work

# Chapter 6

# Conclusions, Social Applications, and Future Work

This final chapter brings together and summarises the main points and important results presented in this thesis with reference to the research objectives of Chapter 1. This thesis is divided into four main parts. *Part* I was focused on the problem statement and main contributions. This Part comprised Chapters 1 and 2. Following this, two experimental parts were described: *Part* II, which have focused on mobile biometric authentication or recognition (Chapters 3 and 4); and *Part* III which has described a case of study in mobile biometric privacy (Chapter 5). Finally, *Part* IV concludes with the thesis, includes social applications of the thesis and provides new horizons that can be explored (Chapter 6).

The **major contributions** made in this thesis are:

- A comprehensive examination has been conducted of the sensors and raw data commonly found in modern mobile devices, with a particular focus on background

sensors. We also provided detailed insights into typical application scenarios and the underlying purposes of data collection in mobile devices. Within the field of subject recognition, we conducted a thorough analysis of the current state of the art in deep learning techniques, specifically addressing the recognition of gait and swipe patterns on mobile devices. Moreover, we extensively examined the extraction of personal and sensitive data from background sensors, encompassing demographics, activity, and behaviour. Lastly, we summarised the metrics proposed in existing literature for assessing privacy concerns related to sensitive data, while providing an overview of methods for safeguarding such information. This contribution helps to raise awareness of the potential privacy risks associated with sensors on mobile devices. By comprehensively examining vulnerabilities, it informs individuals and organisations about the privacy implications of sensor use, enabling them to make implementation decisions and take the necessary precautions. Furthermore, by identifying vulnerabilities, the study contributes to the development of enhance security measures and best practices. Finally, it encourages research and development efforts aimed at creating more secure sensor technologies and better privacy protection mechanisms.

- *Mobile Biometric Authentication*: We have conducted a comprehensive analysis of behavioural biometric authentication or recognition systems on mobile devices specifically focus on gait, swipe, and keystroke, offering a higher level of security compared to previous approaches. To achieve this, we presented an overview of essential concepts related to Transformers, highlighting the distinctions amongst prominent architectures proposed in the literature, such as Vanilla, Informer, Autoformer, Block-Recurrent Transformer, and THAT. Additionally, we explored the potential of Transformers in the field of behavioural biometrics across various modalities on mobile devices. Our research included an extensive experimental analysis of the proposed systems, using popular databases in gait: whuGAIT (Zou

et al., 2020) and OU-ISIR (Iwama et al., 2012; Ngo et al., 2014) databases, swipe: an in-house database, Frank DB (Frank et al., 2012) and HuMIDB (Acien et al., 2021a), and keystroke dynamics: Aalto mobile keystroke database (Palin et al., 2019). In addition, the experimental frameworks developed as part of this research are openly available to the research community, with the aim of advancing the state of the art in gait[1], swipe[2], and keystroke[3] biometrics.

- *Mobile Biometric Privacy*: In our study, we have presented an innovative method called GaitPrivacyON, which employs unsupervised learning for mobile gait biometric verification. This approach not only achieves precise recognition outcomes but also prioritises the preservation of the subject's privacy. Additionally, we have performed a quantitative analysis of GaitPrivacyON using three widely recognised gait recognition databases, MotionSense (Malekzadeh et al., 2018), MobiAct (Vavoulas et al., 2016), and OU-ISIR (Ngo et al., 2014). This contribution also prevents identity theft or fraudulent activities.

## 6.1 Conclusions

The specific findings for each Chapter are described in more detail below.

Chapter 1 initially provided an introduction to fundamental concepts of mobile devices, biometrics, and various modalities and applications of biometric systems. Subsequently, mobile behavioural biometrics have been presented, which is one of the main topics of study in this thesis. Additionally, we presented an overview of the success of DL techniques such as CNNs and RNNs, which in turn inspired the exploration of novel DL architectures undertaken within this thesis. Lastly, privacy preservation for mobile

[1] https://github.com/BiDAlab/ExploringTransformers
[2] https://github.com/BiDAlab/SwipeFormer
[3] https://github.com/BiDAlab/TypeFormer

biometrics has been explored, specifically sensitive data protection. This chapter concluded by explaining the motivation and research contributions originated in this thesis.

Chapter 2 gave an overview of the most relevant mobile background sensors and their main applications. In particular, a comprehensive study on mobile biometric recognition, reviewing the related works of this thesis: *i)* gait identification, *ii)* gait verification, and *iii)* swipe verification. The sensitive data that can be extracted from these sensors has then been described, focusing on demographic, activity, and behavioural data. The chapter concludes with the main privacy-preserving metrics and methods.

*Part II* was the first experimental part, composed of Chapters 3 and 4.

In Chapter 3, we have explored and introduced novel behavioural biometric systems based on Transformers. To the best of our knowledge, this thesis represents the first comprehensive framework for employing Transformers in gait biometrics. The experimental framework involved the evaluation of various Transformer architectures, including Vanilla, Informer, Autoformer, Block-Recurrent Transformer, and THAT, along with a newly proposed configuration: M-GaitFormer. The analysis was conducted using two widely-used public databases, whuGAIT and OU-ISIR. A thorough experimental analysis of the proposed system was performed in two configurations: *i)* identification and *ii)* verification. The main findings of this Chapter are as follows:

- In the identification task, M-GaitFormer outperforms both previous Transformer architectures and traditional deep learning architectures (such as CNNs, RNNs, and CNNs + RNNs) in terms of performance, as assessed on both databases. Remarkably, when applied to the demanding OU-ISIR database, the proposed Transformer achieves a Rank-1 accuracy of 93.33%, demonstrating significant accuracy advancements compared to other techniques. Additionally, the proposed Transformer outperformed existing gait recognition systems in the literature. It

is worth emphasising that M-GaitFormer improved time complexity and memory usage compared to conventional deep learning models.

- In the verification task, M-GaitFormer also outperforms the previous systems in the literature, achieving EER results of 3.42% and 2.92% on the whuGAIT and OU-ISIR databases respectively.

Chapter 4 introduces multiple adaptations of the architecture discussed in Chapter 3. Specifically, we present SwipeFormer, to the best of our knowledge, the first attempt to apply Transformers to touchscreen verification. While sharing certain similarities with M-GaitFormer, this Transformer-based architecture has been modified to accommodate the unique features of touchscreen data. In particular, the Channel Module has been replaced with a Frequency Module. Considering the Frequency Module, the input swipe sequence $X$ is represented in the frequency domain by a discrete Fourier transformation $X_f$. After this, a GRE is included to preserve frequency information. Following an identical architecture to the Temporal Module, the Frequency Module contains a multi-head Self-Attention mechanism and a one-dimensional multi-scale swipe CNN. The main findings of this Chapter are as follows:

- To the best of our knowledge, this is the first study that analyses unconstrained touchscreen gestures, achieving promising results.
- We show how the different data sources (i.e., touchscreen and background sensors) contribute to the system performance and the differences among the two most popular operating systems (i.e., Android and iOS). Under this challenging scenario, SwipeFormer is able to achieve impressive EER values of 6.60% and 3.60% on Android and iOS, respectively, showing that the proposed model is more robust in comparison with recent approaches.

- A validation of the proposed SwipeFormer using the popular publicly available databases collected under constrained conditions: Frank DB (Frank et al., 2012) and HuMIDB (Acien et al., 2020). SwipeFormer achieves EER values of 11.50% and 5.60% on Frank DB and HuMIDB, respectively, outperforming previous state-of-the-art approaches.

- In addition to the use of Transformer-based systems for the gait (Chapter 3) and swipe (Chapter 4) tasks, we have also conducted an assessment of their applicability to mobile keystroke dynamics. TypeFormer consistently outperforms previous approaches across all evaluation scenarios.

Finally, *Part III* contains Chapter 5, focusing on mobile biometric privacy. The data collected from mobile behavioral biometrics for recognition purposes contains personal and sensitive information, such as demographics (e.g., gender, age, ethnicity) and the activities performed by individuals. Therefore, this Chapter introduces GaitPrivacyON, which provides accurate recognition results while preserving subject privacy. The key findings of this chapter are as follows:

- To the best of our knowledge, GaitPrivacyON is the first mobile gait verification approach that incorporates privacy-preserving methods trained in an unsupervised way. It comprises two modules: *i)* two convolutional Autoencoders with shared weights that transform the raw biometric data into a new privacy-preserving representation (e.g., gender or activity), and *ii)* a mobile gait verification system based on a combination of CNNs and RNNs with a Siamese architecture.

- A comprehensive quantitative analysis of GaitPrivacyON is conducted using three popular databases in the field of gait recognition: MotionSense (Malekzadeh et al., 2018), MobiAct (Vavoulas et al., 2016), and OU-ISIR (Ngo et al., 2014). The

results demonstrate accurate verification performance, surpassing 96.60% in terms of the AUC, while effectively reducing the recognition rate of sensitive data to approximately 50% AUC.

## 6.2 Social Applications of the Thesis

The use of biometrics on mobile devices brings society greater security, convenience and improved subject experience. It enables secure recognition, simplifies mobile payments, prevents unauthorised access, and has potential applications in healthcare. This thesis contributes to this field by presenting new biometric recognition models that enhance the existing literature on various behavioural biometrics such as gait, swipe, and keystroke using data extracted from background sensors. Through empirical demonstrations, these techniques illustrate their effectiveness in achieving the aforementioned benefits.

However, it is crucial to strike a balance between the positive aspects of biometrics and the protection of privacy and data security to ensure that the technology is used responsibly and ethically. It is essential to address the potential extraction of sensitive data, as governed by regulations such as the GDPR, which covers personally identifiable information such as gender, ethnicity or political opinion, among others. Existing literature has illustrated the feasibility of extracting such data from the background sensors of mobile devices. Therefore, in response to this challenge, we have developed a behavioural biometric system that serves to secure and protect society against privacy breaches, thereby protecting the sensitive data mentioned above.

## 6.3 Answer from the Research Questions

In this section an explanation of the research questions showed in the introduction are given:

- Do the mobile background sensors sufficiently support behavioural biometrics? Can these sensors be effectively utilised to identify subjects? Certainly. In Chapters 3 and 4 of the thesis, a comprehensive exploration was undertaken to assess the suitability of mobile background sensors for behavioural biometrics. Through rigorous experimentation and analysis, the research findings affirm that these sensors indeed exhibit exceptional performance in accurately capturing and characterising behavioural patterns. The detailed examination and empirical evidence presented in these chapters substantiate the affirmative response, establishing the viability of utilising mobile background sensors for effective subject identification.

- Can these new Transformer architectures outperform previous behavioural biometric approaches such as CNNs and RNNs? Yes, the new Transformer architectures showcased in Chapters 3 and 4 have demonstrated a capability to surpass previous behavioural biometric approaches, including CNNs and RNNs. The research conducted in these chapters establishes that the introduced Transformer architectures have achieved state-of-the-art results, showcasing their superiority in capturing and understanding intricate behavioural patterns for biometric identification.

- Is it possible to achieve a balance between privacy and authentication performance in behavioural biometrics, such as in gait? Yes, achieving a balance between privacy and authentication performance in behavioral biometrics, specifically in gait, is indeed possible. Chapter 5 of the thesis presents a case study where a meticulous approach was implemented. The research demonstrates that it is feasible to

maintain nearly identical recognition performance while concurrently safeguarding sensitive attributes, such as gender and activity. This emphasises the successful accomplishment of a harmonious equilibrium between privacy preservation and recognition accuracy in the context of behavioural biometrics, particularly gait analysis.

## 6.4 Future Work

A number of research lines arise from the work carried out in this thesis. We consider of special interest the following ones:

- Regarding *Part II: Mobile Biometric Authentication*, our proposed Transformers have been analysed using segment-based data. Consequently, to adapt the systems for continuous environments (Papavasileiou et al., 2021), further modifications are necessary. Additionally, future work will explore the potential of the proposed Transformers in other behavioural biometric modalities, such as handwritten signatures (Tolosana et al., 2021b, 2022b), electrocardiograms (Melzi et al., 2022b), voice and speech (Dong et al., 2018), among others. In turn, a multimodal model could be studied, trained with several behavioural biometric traits at the same time or a combination of behavioural and physiological traits. This could lead to: *i)* a two-factor authentication system where multiple biometric features are used to improve overall security, or *ii)* an adaptive authentication system where the most appropriate and reliable biometric feature is dynamically selected and used for recognition depending on the specific context. Moreover, since the proposed Transformer model is based on time series, the models proposed in this thesis can be extrapolated to any other time series domain, such as cloud computing. In addition, as feature work can be try to improve the performance of our

systems considering DL models for the synthesis of data such as Variational Autoencoders (VAEs), Generative Adversarial Network (GAN), or diffusion models. These approaches can significantly improve one- and few-shot learning scenarios as demonstrated in Tolosana et al. (2022b).

- Regarding *Part III: Mobile Biometric Privacy*, the privacy aspects of mobile recognition need to be further addressed, as highlighted in (Delgado-Santos et al., 2022a). One of the main concerns is, that attackers are assumed to have limited resources, but attackers who possess a database of protected biometric data tagged according to soft biometric attributes can also train classifiers in the protected domain. Therefore, it is necessary to study the different types of attacks proposed in the literature and analyse each of them before and after our system, in order to check for weaknesses and improve these aspects.

- In the ensuing phases of our research, we plan to undertake a thorough examination of various attacks in behavioural biometrics. Potential threats encompass adversarial attacks on gait recognition or endeavours to manipulate behavioural biometric data. This inquiry will involve the exploration of innovative approaches and technologies intended to counteract such threats, contributing significantly to the advancement of secure and privacy-preserving behavioural biometric systems. Concurrently, we will conduct a comprehensive investigation into diverse Biometric Privacy-Enhancing Technologies (BPETs) (Melzi et al., 2022a). This exploration will include, but is not limited to, pseudonymisation and encryption techniques. The objective is to identify the most effective solutions for addressing our privacy concerns. Furthermore, we intend to generate new synthetic data to be used for both training and testing these systems, with the aim of enhancing privacy and recognition outcomes. This dual-focused study aims to fortify the robustness and privacy safeguards within behavioural biometric systems.

# Appendix A

# Resumen Extendido de la Tesis

**Aproximaciones Disruptivas para la Mejora de la Autenticación y Privacidad de los Sistemas Biométricos Conductuales en Escenarios Móviles**

## A.1 Resumen

El creciente número de dispositivos móviles en los últimos años ha ocasionado la recopilación de una gran cantidad de información personal que necesita ser protegida. Con este fin, la biometría conductual se ha vuelto muy popular. Pero, ¿cuál es el poder discriminativo de la biometría conductual en dispositivos móviles?

Con el éxito del *Deep Learning* (DL), las arquitecturas basadas en *Convolutional Neural Networks* (CNNs) y *Recurrent Neural Networks* (RNNs), como *Long Short-Term Memory* (LSTM), han mostrado mejoras en rendimiento y robustez en comparación con los métodos tradicionales de *machine learning*. Sin embargo, estas arquitecturas de DL, aunque han sido ampliamente utilizadas, siguen presentando ciertas limitaciones que es

necesario abordar. Para resolver estos problemas, han surgido nuevas arquitecturas de DL como los Transformers. La pregunta es, ¿pueden los Transformers ser empleados en biometría y mejorar el rendimiento del estado del arte?

Como una forma de encontrar las respuestas a estas preguntas, esta tesis analiza la autenticación biométrica conductual con datos adquiridos de los sensores *background* en dispositivos móviles (es decir, acelerómetros y giroscopios). Además, esta es la primera tesis que explora y propone nuevos sistemas biométricos conductuales basados en Transformers, con el fin de superar el rendimiento del estado del arte en la forma de caminar, la interacción táctil (*swipe*) y la dinámica del tecleo.

El uso de la biometría requiere un delicado equilibrio entre seguridad y privacidad. Las modalidades biométricas proporcionan un enfoque único e inherentemente personal para la autenticación. No obstante, la biometría también pueden conllevar una invasión de la privacidad personal. Según la *General Data Protection Regulation* (GDPR) introducida por la Unión Europea, los datos personales, como los biométricos, son datos sensibles y deben utilizarse y protegerse adecuadamente. Esta tesis analiza el impacto de los datos sensibles en el rendimiento de los sistemas biométricos y propone un novedoso enfoque no supervisado que preserva la privacidad.

La investigación llevada a cabo en esta tesis ha dado lugar a importantes contribuciones, entre ellas: *i)* una revisión exhaustiva de las vulnerabilidades de privacidad de los sensores de dispositivos móviles, que abarca una revisión de las métricas propuestas en la literatura para cuantificar la privacidad en relación con los datos sensibles, junto con una visión general de los métodos de protección para proteger la información sensible; *ii)* un análisis de los sistemas de autenticación por biometría conductual en dispositivos móviles (la forma de caminar, la interacción táctil (*swipe*) y la dinámica del tecleo), siendo la primera tesis que explora el potencial de los Transformers para la biometría

conductual, presentando arquitecturas novedosas que mejoran el estado del arte; y *iii)* un nuevo enfoque que proporciona resultados precisos de autenticación al tiempo que preserva la privacidad del sujeto.

## A.2 Conclusiones

Los hallazgos específicos de cada capítulo se describen con más detalle a continuación.

El Capítulo 1 proporciona una introducción inicial a los conceptos fundamentales de los dispositivos móviles, la biometría y las diversas modalidades y aplicaciones de los sistemas biométricos. Posteriormente, se presentan los aspectos de la biometría conductual en dispositivos móviles, tema abordado dentro de la presente tesis. Además, se presenta una visión general del éxito de las técnicas de DL como CNN y RNN, que han servido para realizar una exploración de nuevas arquitecturas de DL desarrolladas en esta tesis. Por último, se analiza el tema de la privacidad para la biometría en dispositivos móviles, específicamente la protección de datos sensibles. Este capítulo concluye explicando la motivación y las contribuciones de investigación originadas en esta tesis.

El Capítulo 2 proporciona una visión general de los sensores *background* en dispositivos móviles más relevantes y sus principales aplicaciones. En particular, se realiza un estudio exhaustivo sobre la autenticación biométrica en dispositivos móviles, revisando los trabajos relacionados de esta tesis: *i)* la identificación y verificación de la forma de caminar y *ii)* la verificación a partir de la interacción táctil (*swipe biometrics*) realizada sobre la pantalla del dispositivo móvil. Tras ello, se describen los datos sensibles que se pueden extraer de estos sensores, centrándose en datos demográficos, de actividad y de comportamiento. El capítulo concluye con las principales métricas y métodos de

preservación de la privacidad.

La *Parte II* de la presente tesis constituye la primera parte experimental, compuesta por los Capítulos 3 y 4, centrada en el análisis de los sistemas biométricos conductuales desde el punto de vista de la autenticación.

En el Capítulo 3, se exploran nuevos sistemas biométricos conductuales basados en Transformers. Esta tesis representa el primer marco integral que propone con éxito nuevas aproximaciones basadas en Transformers para la autenticación de la persona a través de la forma de caminar. El marco experimental se compone de varias arquitecturas de Transformers, incluyendo *Vanilla*, *Informer*, *Autoformer*, *Block-Recurrent Transformer* y *THAT*, junto con una nueva configuración propuesta: *M-GaitFormer*. El análisis se realiza utilizando dos bases de datos públicas ampliamente utilizadas, *whuGAIT* y *OU-ISIR*. Se realiza un análisis experimental del sistema propuesto en dos configuraciones: *i)* identificación y *ii)* verificación. Los principales hallazgos de este capítulo son los siguientes:

- En la tarea de identificación, *M-GaitFormer* supera tanto a las arquitecturas de Transformers anteriores como a las arquitecturas tradicionales de DL (CNN, RNN y CNN + RNN) en términos de rendimiento, evaluado en ambas bases de datos. Notablemente, cuando se aplica a la exigente base de datos *OU-ISIR*, el Transformer propuesto logra una precisión de clasificación en *Rank-1* del 93.33%, demostrando avances significativos en comparación con otras técnicas. Además, el Transformer propuesto supera en rendimiento a los sistemas en el estado del arte. Es importante destacar que M-GaitFormer mejora el tiempo computacional y el uso de memoria en comparación con los modelos convencionales de DL.

- En la tarea de verificación, *M-GaitFormer* también supera a los sistemas anteriores en la literatura, logrando resultados en términos de *Equal Error Rate* (EER) del

3.42% y 2.92% en las bases de datos *whuGAIT* y *OU-ISIR*, respectivamente.

El Capítulo 4 introduce múltiples adaptaciones de la arquitectura discutida en el Capítulo 3 para los sistemas biométricos basados en la interacción táctil (*swipe*) con la pantalla de los dispositivos móviles. Específicamente, se presenta *SwipeFormer*, la primera aproximación en la literatura basada en Transformers. Si bien comparte ciertas similitudes con *M-GaitFormer*, esta arquitectura basada en Transformers se ha modificado para adaptarse a las características únicas de los datos táctiles. En particular, el *Channel Module* se ha reemplazado por un *Frequency Module*, extrayendo por tanto información temporal y frecuencial haciendo uso de la Transformada Discreta de Fourier. Después de esto, se incluye un módulo llamado *Gaussian Range Encoding* (GRE) para preservar la información de frecuencia. Siguiendo una arquitectura idéntica al *Temporal Module*, el *Frequency Module* contiene un módulo de *multi-head Self-Attention mechanism* y una *one-dimensional multi-scale swipe CNN*. Los principales hallazgos de este capítulo son los siguientes:

- La presente Tesis supone que este es el primer estudio que analiza gestos táctiles (*swipe*) en escenarios de aplicación no controlados, logrando resultados prometedores.
- Se analiza cómo las diferentes fuentes de datos (es decir, táctil y sensores *background*) contribuyen al rendimiento del sistema y las diferencias entre los dos sistemas operativos más populares (es decir, *Android* e *iOS*). En este escenario desafiante, *SwipeFormer* logra valores impresionantes de *EER* del 6.60% y 3.60% en *Android* e *iOS*, respectivamente, lo que demuestra que el modelo propuesto es más robusto en comparación con enfoques recientes.
- Se realiza una validación de *SwipeFormer* utilizando bases de datos públicas disponibles en la literatura: *Frank DB* (Frank et al., 2012) y *HuMIDB* (Acien et al., 2020).

*SwipeFormer* logra valores de *EER* del 11.50% y 5.60% en *Frank DB* y *HuMIDB*, respectivamente, superando en rendimiento de otras aproximaciones en el estado del arte.

- Además del uso de sistemas basados en Transformers para las tareas de la forma de caminar (Capítulo 3) y e interacción táctil (Capítulo 4), también se realiza una evaluación de su aplicabilidad a la dinámica del tecleo en dispositivos móviles. *TypeFormer* supera consistentemente en rendimiento a enfoques anteriores en todos los escenarios de evaluación.

Finalmente, la *Parte III* contiene el Capítulo 5, centrado en la privacidad biométrica en escenarios móviles. Los datos recopilados de la biometría conductual en escenarios móviles con fines de autenticación contienen información personal y sensible, como datos demográficos (por ejemplo, género, edad y etnia) y las actividades realizadas por las personas. Por lo tanto, este capítulo presenta *GaitPrivacyON*, una aproximación que proporciona resultados precisos de autenticación al tiempo que preserva la privacidad del sujeto. Los principales hallazgos de este capítulo son los siguientes:

- *GaitPrivacyON* es el primer enfoque de verificación la forma de caminar usando dispositivos móviles que incorpora métodos de preservación de la privacidad entrenados de manera no supervisada. Consta de dos módulos: *i)* dos *autoencoders* convolucionales con pesos compartidos que transforman los datos biométricos en bruto en una nueva representación más segura en términos de privacidad (por ejemplo, género o actividad), y *ii)* un sistema de verificación de la forma de caminar basado en una combinación de CNNs y RNNs con una arquitectura siamesa.
- Se realiza un análisis cuantitativo de GaitPrivacyON utilizando tres bases de datos populares en el ámbito de la forma de caminar con dispositivos móviles: MotionSense (Malekzadeh et al., 2018), MobiAct (Vavoulas et al., 2016) y OU-ISIR (Ngo

et al., 2014). Los resultados demuestran un rendimiento preciso de verificación, superando el 96.60% en términos del *Area Under the Receiver Operating Characteristic* (AUC), al tiempo que se reduce efectivamente los datos sensibles (genero o actividad realizada por el sujeto) a aproximadamente el 50% AUC.

## A.3 Aplicaciones Sociales de la Tesis

El uso de la biometría en dispositivos móviles brinda a la sociedad una mayor seguridad, comodidad y una experiencia mejorada para los sujetos. Permite una autenticación segura, simplifica los pagos móviles, previene el acceso no autorizado y tiene aplicaciones potenciales en el ámbito de la salud. Esta tesis contribuye a este campo al presentar nuevos modelos de autenticación biométrica que mejoran la literatura existente sobre diversas biometrías conductuales, como la forma de caminar, la interacción táctil (*swipe*) y la dinámica del tecleo, utilizando datos extraídos de los sensores *background*. A través de demostraciones empíricas, estas técnicas ilustran su eficacia para lograr los beneficios mencionados anteriormente.

Sin embargo, es crucial encontrar un equilibrio entre los aspectos positivos de la biometría y la protección de la privacidad y la seguridad de los datos, para garantizar que la tecnología se utilice de manera responsable y ética. Es esencial abordar la posible extracción de datos sensibles, regulados por normativas como el GDPR, que cubre información personal identificable, como género, etnia u opinión política, entre otros. La literatura existente ha demostrado la viabilidad de extraer dichos datos de los sensores *background* de los dispositivos móviles. Por lo tanto, como respuesta a este desafío, se han desarrollado sistemas biométricos conductuales que sirvan para asegurar y proteger a la sociedad contra violaciones de privacidad, protegiendo así los datos sensibles mencionados anteriormente.

## A.4 Líneas de Trabajo Futuro

De la labor realizada en esta tesis se derivan varias líneas de investigación que consideramos de especial interés:

- En cuanto a la *Parte II: Autenticación Biométrica Móvil*, nuestros Transformers propuestos han sido analizados utilizando datos obtenidos de forma puntual. Por lo tanto, para adaptar los sistemas a entornos continuos (Papavasileiou et al., 2021), se necesitan modificaciones adicionales. Además, trabajos futuros explorarán el potencial de los Transformers propuestos en otras modalidades biométricas conductuales, como firmas manuscritas (Tolosana et al., 2021b, 2022b), electrocardiogramas (Melzi et al., 2022b), voz y habla (Dong et al., 2018), entre otras. A su vez, se podría estudiar un sistema multimodal, entrenado con varios rasgos biométricos conductuales al mismo tiempo o una combinación de rasgos conductuales y fisiológicos. Esto podría llevar a: *i)* un sistema de autenticación de dos factores donde se utilizan múltiples características biométricas para mejorar la seguridad general, o *ii)* un sistema de autenticación adaptativa donde se selecciona y utiliza dinámicamente la característica biométrica más adecuada y confiable para la autenticación según el contexto específico. Además, dado que el modelo Transformer propuesto se basa en secuencias temporales, los modelos propuestos en esta tesis se pueden extrapolar a cualquier otro dominio de series temporales, como la computación en la nube. Además, se puede intentar mejorar el rendimiento de nuestros sistemas para la síntesis de datos considerando modelos DL como *Variational Autoencoders* (VAEs), *Generative Adversarial Network* (GAN) o *diffusion models*. Estos enfoques pueden mejorar significativamente los escenarios de aprendizaje de *one-shot* o *few-shot learning*, como se demuestra en Tolosana et al. (2022b).

- En relación a la *Parte III: Privacidad Biométrica en Dispositivos Móviles*, es necesario continuar investigando los aspectos de privacidad relacionados con la autenticación móvil, tal como se resalta en (Delgado-Santos et al., 2022a). Una de las principales preocupaciones radica en que se asume que los atacantes tienen recursos limitados. Sin embargo, los atacantes puedes poseer datos etiquetados con información personal, teniendo la posibilidad de entrenar nuevos clasificadores. Por lo tanto, es necesario estudiar los diferentes tipos de ataques propuestos en la literatura y analizar cada uno de ellos antes y después de la implementación de nuestro sistema, con el fin de detectar debilidades y mejorar estos aspectos. Simultáneamente, se debe llevar a cabo una investigación exhaustiva que abarque diversas *Biometric Privacy-Enhancing Technologies* (BPETs) (Melzi et al., 2022a), incluyendo, pero no limitándose a la pseudonimización y el cifrado, para determinar la solución óptima para abordar nuestro problema. A su vez, se propone como trabajo futuro la creación de nuevos datos sintéticos tanto para entrenar como para atacar a estos sistemas. Con ello se pretende mejorar la privacidad y los resultados en la tarea de autenticación.

# References


EU 2016/679 (General Data Protection Regulation). https://gdpr-info.eu/, 2016.

Health Informatics — Pseudonymization. Technical report, International Organization for Standardization, 2017.

PriMa: Privacy Matters, H2020-MSCA-ITN-2019-860315. https://www.prima-itn.eu/, 2019.

TReSPAsS-ETN: TRaining in Secure and PrivAcy-preserving biometricS, H2020-MSCA-ITN-2019-860813. https://www.trespass-etn.eu/, 2019.

Number of Apps Available in Leading App Stores as of 3rd Quarter 2020. https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/, 2021.

Number of Smartphone Subscriptions Worldwide from 2016 to 2021, with Forecasts from 2022 to 2027. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/, 2023. Accessed: 2023-03-17.

M. Abadi, A. Chu, I. Goodfellow, H. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep Learning with Differential Privacy. In *Proc. ACM SIGSAC Conference on Computer and Communications Security*, 2016.

M. Abuhamad, A. Abusnaina, D. Nyang, and D. Mohaisen. Sensor-Based Continuous Authentication of Smartphones' Users Using Behavioral Biometrics: A Contemporary Survey. *IEEE Internet of Things Journal*, 8(1):65–84, 2021.

A. Acien, A. Morales, J. Fierrez, R. Vera-Rodriguez, and J. Hernandez-Ortega. Active Detection of Age Groups Based on Touch Interaction. *IET Biometrics*, 8(1):101–108, 2019a.

A. Acien, A. Morales, R. Vera-Rodriguez, J. Fierrez, and R. Tolosana. Multilock: Mobile Active Authentication based on Multiple Biometric and Behavioral Patterns. In *Proc. International Workshop on Multimodal Understanding and Learning for Embodied Applications*, 2019b.

A. Acien, A. Morales, R. Vera-Rodriguez, and J. Fierrez. Smartphone Sensors for Modeling Human-Computer Interaction: General Outlook and Research Datasets for User Authentication. In *Proc. IEEE Annual Computers, Software and Applications Conference*, 2020.

A. Acien, A. Morales, J. Fierrez, R. Vera-Rodriguez, and O. Delgado-Mohatar. BeCAPTCHA: Behavioral Bot Detection using Touchscreen and Mobile Sensors benchmarked on HuMIdb. *Engineering Applications of Artificial Intelligence*, 98:104058, 2021a.

A. Acien, A. Morales, J. V. Monaco, R. Vera-Rodriguez, and J. Fierrez. TypeNet: Deep Learning Keystroke Biometrics. *IEEE Transactions on Biometrics, Behavior and Identity Science*, 4(1):57–70, 2021b.

C. C. Aggarwal. On K-Anonymity and the Curse of Dimensionality. In *Proc. International Conference on Very Large Data Bases*, 2005.

C. C. Aggarwal and P. S. Yu. *A Survey of Randomization Methods for Privacy-Preserving Data Mining*. Citeseer, 2008.

D. Agrawal and C. Aggarwal. On the Design and Quantification of Privacy Preserving Data Mining Algorithms. In *Proc. ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2001.

R. Agrawal and R. Srikant. Privacy-preserving Data Mining. In *Proc. ACM SIGMOD International Conference on Management of Data*, 2000.

R. Agrawal, T. Imieliński, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. In *Proc. ACM SIGMOD International Conference on Management of Data*, 1993.

A. Aljeraisy, M. Barati, O. Rana, and C. Perera. Privacy Laws and Privacy by Design Schemes for the Internet of Things: A Developer's Perspective. *ACM Computing Surveys*, 54(5), May 2021.

A. Almaatouq, P. rieto Castrillo, and A. Pentland. Mobile Communication Signatures of Unemployment. In *Proc. International Conference on Social Informatics*, 2016.

F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Gonzalez-Rodriguez. Quality-based Conditional Processing in Multi-Biometrics: Application to Sensor Interoperability. *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans*, 40(6):1168–1179, 2010.

M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *Proc. ACM SIGSAC Conference on Computer & Communications Security*, 2013.

A. Anjum and M. U. Ilyas. Activity Recognition using Smartphone Sensors. In *Proc. IEEE Consumer Communications and Networking Conference*, 2013.

M. Antal, Z. Bokor, and L. Z. Szabó. Information Revealed from Scrolling Interactions on Mobile Devices. *Pattern Recognition Letters*, 56:7–13, 2015.

A. D. Antar, M. Ahmed, and M. Ahad. Challenges in Sensor-based Human Activity Recognition and a Comparative Analysis of Benchmark Datasets: a Review. In *Proc. International Conference on Informatics, Electronics & Vision and International Conference on Imaging, Vision & Pattern Recognition*, 2019.

C. A. Ardagna, M. Cremonini, E. Damiani, S. D. Vimercati, and P. Samarati. Location Privacy Protection through Obfuscation-based Techniques. In *Proc. IFIP Annual Conference on Data and Applications Security and Privacy*, 2007.

Z. Arnold, D. Larose, and E. Agu. Smartphone Inference of Alcohol Consumption Levels from Gait. In *Proc. International Conference on Healthcare Informatics*, 2015.

S. Barth, M. D. de Jong, M. Junger, P. H. Hartel, and J. C. Roppelt. Putting the Privacy Paradox to the Test: Online Privacy and Security Behaviors among Users with Technical Knowledge, Privacy Awareness and Financial Resources. *Telematics and Informatics*, 41:55–69, 2019.

I. Beltagy, M. E. Peters, and A. Cohan. Longformer: The Long-Document Transformer. *arXiv preprint arXiv:2004.05150*, 2020.

K. Bhaduri, M. D. Stefanski, and A. N. Srivastava. Privacy-Preserving Outlier Detection Through Random Nonlinear Data Distortion. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 41(1):260–272, 2011.

C. Bo, L. Zhang, T. Jung, J. Han, X.-Y. Li, and Y. Wang. Continuous User Identification via Touch and Movement Behavioral Biometrics. In *Proc. IEEE International Performance Computing and Communications Conference*, 2014.

M. Boakes, R. Guest, F. Deravi, and B. Corsetti. Exploring Mobile Biometric Performance through Identification of Core Factors and Relationships. *IEEE Transactions on Biometrics, Behavior and Identity Science*, 1(4):278–291, 2019.

B. Bortolato, M. Ivanovska, P. Rot, J. Križaj, P. Terhörst, N. Damer, P. Peer, and V. Štruc. Learning Privacy-enhancing Face Representations through Feature Disentanglement. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2020.

A. Boutet, C. Frindel, S. Gambs, T. Jourdan, and R. C. Ngueveu. DYSAN: Dynamically Sanitizing Motion Sensor Data against Sensitive Inferences through Adversarial Networks. In *Proc. ACM Asia Conference on Computer and Communications Security*, 2021.

K. Bowyerin and M. Burge. *Handbook of Iris Recognition.* Springer, 2016.

R. Brand. Microdata Protection through Noise Addition. *Inference Control in Statistical Databases: From Theory to Practice*, pages 97–116, 2002.

J. Burgues, J. Fierrez, D. Ramos, M. Puertas, and J. Ortega-Garcia. Detecting Invalid Samples in Hand Geometry Verification Through Geometric Measurements. In *Proc. IAPR International Conference on Pattern Recognition*, pages 113–120, 2010.

J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. Srivastava. Participatory Sensing. *UCLA: Center for Embedded Network Sensing*, 2006.

M. Chamikara, P. Bertok, D. Liu, S. Camtepe, and I. Khalil. Efficient Data Perturbation for Privacy Preserving and Accurate Data Stream Mining. *Pervasive and Mobile Computing*, 48:1–19, 2018.

C. H. Chan, X. Zou, N. Poh, and J. Kittler. Illumination Invariant Face Recognition: A Survey. In *Computer Vision: Concepts, Methodologies, Tools, and Applications*, pages 58–79. IGI Global, 2018.

L. Chang, J. Lu, J. Wang, X. Chen, D. Fang, Z. Tang, P. Nurmi, and Z. Wang. Sleep-

Guard: Capturing Rich Sleep Information using Smartwatch Sensing Data. *Proc. in ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–34, 2018.

K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi. Broadening the Scope of Differential Privacy using Metrics. In *Proc. International Symposium on Privacy Enhancing Technologies Symposium*, 2013.

K. Chaudhuri and N. Mishra. When Random Sampling Preserves Privacy. In *Proc. Advances in Cryptology*, 2006.

R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, J. d. R. M. G. Tröster, and D. Roggen. The opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognition Letters*, 34(15):2033–2042, 2013.

S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee. Toward Privacy in Public Databases. In *Proc. Theory of Cryptography Conference*, 2005.

K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu. Deep Learning for Sensor-Based Human Activity Recognition: Overview, Challenges and Opportunities. *ACM Computing Surveys*, 54(4), 2021.

L. Chen, Z. Mu, B. Zhang, and Y. Zhang. Ear Recognition from One Sample per Person. *PloS one*, 10(5):e0129505, 2015.

Y. Chen, B. Zheng, Z. Zhang, Q. Wang, C. Shen, and Q. Zhang. Deep Learning on Mobile and Embedded Devices: State-of-the-Art, Challenges and Future Directions. *ACM Computing Surveys*, 53(4), 2020.

Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell. Unobtrusive Sleep Monitoring using Smartphones. In *Proc. International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, 2013.

Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui. WiFi CSI based Passive Human Activity Recognition using Attention based BLSTM. *IEEE Transactions on Mobile Computing*, 18(11):2714–2724, 2018.

R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar. Preserving User Location Privacy in Mobile Data Management Infrastructures. In *Proc. International Workshop on Privacy Enhancing Technologies*, 2006.

L. Chi and X. Zhu. Hashing Techniques: A Survey and Taxonomy. *ACM Computing Surveys*, 50(1), 2017.

R. Child, S. Gray, A. Radford, and I. Sutskever. Generating Long Sequences with Sparse Transformers. *arXiv preprint arXiv:1904.10509*, 2019.

R. S. Choraś. Retina Recognition for Biometrics. In *Proc. International Conference on Digital Information Management*, pages 177–180, 2012.

T. Dalenius. Finding a Needle in a Haystack or Identifying Anonymous Census Records. *Journal of Official Statistics*, 2(3):329, 1986.

L. N. Darlow and B. Rosman. Fingerprint Minutiae Extraction using Deep Learning. In *Proc. International Joint Conference on Biometrics*, pages 22–30, 2017.

A. Das, C. Galdi, H. Han, R. Ramachandra, J.-L. Dugelay, and A. Dantcheva. Recent Advances in Biometric Technology for Mobile Bevices. In *Proc. IEEE International Conference on Biometrics Theory, Applications and Systems*, 2018.

D. Dasgupta, A. Roy, A. Nag, et al. *Advances in User Authentication*. Springer, 2017.

E. Davarci, B. Soysal, I. Erguler, S. O. Aydin, O. Dincer, and E. Anarim. Age Group Detection using Smartphone Motion Sensors. In *Proc. European Signal Processing Conference*, 2017.

K. David and H. Berndt. 6G Vision and Requirements: Is There Any Need for Beyond 5G? *IEEE Vehicular Technology Magazine*, 13(3):72–80, 2018.

L. De Luisa, G. E. Hine, E. Maiorana, and P. Campisi. EEG-based Biometric Verification using Siamese CNNs. In *Proc. International Conference on Image Analysis and Processing*, 2019.

K. Delac and M. Grgic. A Survey of Biometric Recognition Methods. In *Proc. International Symposium on Electronics in Marine*, 2004.

R. Delgado-Escaño, F. M. Castro, J. R. Cózar, M. J. Marín-Jiménez, and N. Guil. An End-to-End Multi-Task and Fusion CNN for Inertial-Based Gait Recognition. *IEEE Access*, 7:1897–1908, 2018.

P. Delgado-Santos, G. Stragapede, R. Tolosana, R. Guest, F. Deravi, and R. Vera-Rodriguez. A Survey of Privacy Vulnerabilities of Mobile Device Sensors. *ACM Computing Surveys*, 54(11), 2022a.

P. Delgado-Santos, R. Tolosana, R. Guest, R. Vera, F. Deravi, and A. Morales. GaitPrivacyON: Privacy-Preserving Mobile Gait Biometrics using Unsupervised Learning. *Pattern Recognition Letters*, 161:30–37, 2022b.

P. Delgado-Santos, R. Tolosana, R. Guest, F. Deravi, and R. Vera-Rodriguez. Exploring Transformers for Behavioural Biometrics: A Case Study in Gait Recognition. *Pattern Recognition*, 2023a.

P. Delgado-Santos, R. Tolosana, R. Guest, P. Lamb, K. andrei, C. Coughlan, and R. Vera-Rodriguez. SwipeFormer: Transformers for Mobile Touchscreen Biometrics. *Expert Systems with Applications*, 2023b.

P. Delgado-Santos, R. Tolosana, R. Guest, R. Vera-Rodriguez, and J. Fierrez. M-

GaitFormer: Mobile Biometric Gait Verification using Transformers. *Engineering Applications of Artificial Intelligence*, 2023c.

V. Dhakal, A. M. Feit, P. O. Kristensson, and A. Oulasvirta. Observations on Typing from 136 Million Keystrokes. In *Proc. CHI Conference on Human Factors in Computing Systems*, 2018.

C. Ding and D. Tao. Pose-Invariant Face Recognition with Homography-Based Normalization. *Pattern Recognition*, 66:144–152, 2017.

J. Domingo-Ferrer and J. Soria-Comas. From t-Closeness to Differential Privacy and Vice Versa in Data Anonymization. *Knowledge-Based Systems*, 74:151–158, 2015.

L. Dong, S. Xu, and B. Xu. Speech-Transformer: a No-Recurrence Sequence-to-Sequence Model for Speech Recognition. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5884–5888. IEEE, 2018.

F. du Pin Calmon and N. Fawaz. Privacy Against Statistical Inference. In *Proc. Allerton Conference on Communication, Control and Computing*. IEEE, 2012.

C. Dwork and A. Roth. The Algorithmic Foundations of Differential Privacy. *Found. Trends® Theor. Computing Sci.*, 9(3-4):211–407, 2014.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, Ourselves: Privacy Via Distributed Noise Generation. In *Proc. International Conference on the Theory and Applications of Cryptographic Techniques*, 2006.

E. Ellavarason, R. Guest, F. Deravi, R. Sanchez-Riello, and B. Corsetti. Touch-Dynamics based Behavioural Biometrics on Mobile Devices–A Review from a Usability and Performance Perspective. *ACM Computing Surveys*, 53(6):1–36, 2020.

A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke. Privacy Preserving Mining of Association Rules. *Information Systems*, 29(4):343–364, 2004.

M. Faundez-Zanuy, J. Fierrez, M. A. Ferrer, M. Diaz, R. Tolosana, and R. Plamondon. Handwriting Biometrics: Applications and Future Trends in E-security and E-health. *Cognitive Computation*, 12(5):940–953, 2020.

T. Feng, J. Yang, Z. Yan, E. M. Tapia, and W. Shi. Tips: Context-aware Implicit User Identification using Touch Screen in Uncontrolled Environments. In *Proc. Workshop on Mobile Computing Systems and Applications*, 2014.

P. Fernandez-Lopez, J. Liu-Jimenez, K. Kiyokawa, Y. Wu, and R. Sanchez-Reillo. Recurrent Neural Network for Inertial Gait User Recognition in Smartphones. *Sensors*, 19(18):1–16, 2019.

J. Fierrez, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Bayesian Adaptation for User-dependent Multimodal Biometric Authentication. *Pattern Recognition*, 38(8):1317–1319, August 2005.

J. Fierrez, A. Morales, R. Vera-Rodriguez, and D. Camacho. Multiple Classifiers in Biometrics. Part 2: Trends and Challenges. *Information Fusion*, 44:103–112, November 2018a.

J. Fierrez, A. Pozo, M. Martinez-Diaz, J. Galbally, and A. Morales. Benchmarking Touchscreen Biometrics for Mobile Authentication. *IEEE Transactions on Information Forensics and Security*, 13(11):2720–2733, 2018b.

C. Filipi Gonçalves dos Santos, D. d. S. Oliveira, L. A. Passos, R. Gonçalves Pires, D. Felipe Silva Santos, L. Pascotti Valem, T. P. Moreira, M. Cleison S. Santana, M. Roder, J. Paulo Papa, et al. Gait Recognition Based on Deep Learning: A Survey. *ACM Computing Surveys*, 55(2):1–34, 2022.

A. I. Filippov, A. V. Iuzbashev, and A. S. Kurnev. User Authentication via Touch Pattern Recognition based on Isolation Forest. In *Proc. IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering*, 2018.

P. Franček, K. Jambrošić, M. Horvat, and V. Planinec. The Performance of Inertial Measurement Unit Sensors on Various Hardware Platforms for Binaural Head-Tracking Applications. *Sensors*, 23(2):872, 2023.

M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. 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, 2012.

M. Gadaleta and M. Rossi. IDNet: Smartphone-based Gait Recognition with Convolutional Neural Networks. *Pattern Recognition*, 74:25–37, 2018.

S. L. Garfinkel. De-identification of Personal Information. *National Institute of Standards and Technology*, 2015.

G. Garofalo, D. Preuveneers, and W. Joosen. Data Privatizer for Biometric Applications and Online Identity Management. *Privacy and Identity Management*, pages 209–225, 2019.

M. Gavrilova, F. Ahmed, S. Azam, P. P. Paul, W. Rahman, M. Sultana, and F. T. Zohra. Emerging Trends in Security System Design using the Concept of Social Behavioural Biometrics. *Information Fusion for Cyber-Security Analytics*, pages 229–251, 2017.

O. Ghahabi and J. Hernando. Deep Learning Backend for Single and Multisession i-vector Speaker Recognition. *Transactions on Audio, Speech and Language Processing*, 25(4):807–817, 2017.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

F. Gravenhorst, A. Muaremi, J. Bardram, A. Grünerbl, O. Mayora, G. Wurzer, M. Frost, V. Osmani, B. Arnrich, P. Lukowicz, et al. Mobile Phones as Medical Devices in Mental Disorder Treatment: an Overview. *Personal and Ubiquitous Computing*, 19 (2):335–353, 2015.

J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, et al. Recent Advances in Convolutional Neural Networks. *Pattern Recognition*, 77:354–377, 2018.

M. Gupta and B. Tripathy. Artificial Intelligence-Based Behavioral Biometrics. *Encyclopedia of Data Science and Machine Learning*, pages 887–898, 2023.

B. Hadjkacem, W. Ayedi, M. B. Ayed, S. A. Alshaya, and M. Abid. A Novel Gait-Appearance-Based Multi-Scale Video Covariance Approach for Pedestrian (Re)-Identification. *Engineering Applications of Artificial Intelligence*, 91:103566, 2020.

O. Hajihassnai, O. Ardakanian, and H. Khazaei. ObscureNet: Learning Attribute-invariant Latent Representation for Anonymizing Sensor Data. In *Proc. International Conference on Internet-of-Things Design and Implementation*, 2021.

M. Haris, H. Haddadi, and P. Hui. Privacy Leakage in Mobile Computing: Tools, Methods and Characteristics. *arXiv preprint arXiv:1410.4978*, 2014.

A. Hasan, Q. Jiang, J. Luo, C. Li, and L. Chen. An Effective Value Swapping Method for Privacy Preserving Data Publishing. *Security and Communication Networks*, 9 (16):3219–3228, 2016.

L. Hernández-Álvarez, J. M. de Fuentes, L. González-Manzano, and L. H. Encinas. Privacy-Preserving Sensor-Based Continuous Authentication and User Profiling: A Review. *Sensors*, 21(1):92, 2021.

B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Preserving Privacy in GPS Traces via Uncertainty-aware Path Cloaking. In *Proc. ACM Conference on Computer and Communications Security*, 2007.

D. Hutchins, I. Schlag, Y. Wu, E. Dyer, and B. Neyshabur. Block-Recurrent Transformers. In *Proc. Advances in Neural Information Processing Systems*, 2022.

H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi. The OU-ISIR Gait Database comprising the Large Population Dataset and Performance Evaluation of Gait Recognition. *IEEE Transactions on Information Forensics and Security*, 7(5):1511–1521, 2012.

Y. Iwasawa, K. Nakayama, I. Yairi, and Y. Matsuo. Privacy Issues Regarding the Application of DNNs to Activity-Recognition using Wearables and Its Countermeasures by Use of Adversarial Training. In *Proc. International Joint Conference on Artificial Intelligence*, 2017.

A. Jain and V. Kanhangad. Investigating Gender Recognition in Smartphones using Accelerometer and Gyroscope Sensor Readings. In *Proc. International Conference on Computational Techniques in Information and Communication Technologies*, 2016.

A. Jain and V. Kanhangad. Gender Recognition in Smartphones using Touchscreen Gestures. *Pattern Recognition Letters*, 125:604–611, 2019.

A. K. Jain, A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004.

A. K. Jain, P. Flynnand, and A. A. Ross. *Handbook of Biometrics*. Springer Science & Business Media, 2007.

A. K. Jain, K. Nandakumar, and A. Ross. 50 Years of Biometric Research: Accomplishments, Challenges and Opportunities. *Pattern Recognition Letters*, 79:80–105, 2016.

F. John Dian, R. Vahidnia, and A. Rahmati. Wearables and the Internet of Things (IoT), Applications, Opportunities and Challenges: A Survey. *IEEE Access*, 8:69200–69211, 2020.

J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual Losses for Real-time Style Transfer and Super-resolution. In *Proc. European Conference on Computer Vision*, 2016.

F. Julien, M. Raya, M. Felegyhazi, and P. Papadimitratos. Mix-Zones for Location Privacy in Vehicular Networks. In *Proc. ACM Workshop on Wireless Networking for Intelligent Transportation Systems*, 2007.

K. Chen and L. Liu. Privacy Preserving Data Classification with Rotation Perturbation. In *Proc. International Conference on Data Mining*, 2005.

M. Kantarcioğlu, J. Jin, and C. Clifton. When Do Data Mining Results Violate Privacy? In *Proc. CM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004.

A. Karakasidis, G. Koloniari, and V. Verykios. Privacy Preserving Blocking and Meta-Blocking. In *Machine Learning and Knowledge Discovery in Databases*, 2015.

M. Kearns, M. Pai, A. Roth, and J. Ullman. Mechanism Design in Large Games: Incentives and Privacy. In *Proc. Conference on Innovations in Theoretical Computer Science*, 2014.

S. Khan, S. Parkinson, L. Grant, N. Liu, and S. Mcguire. Biometric Systems Utilising Health Data from Wearable Devices: Applications and Future Challenges in Computer Security. *ACM Computing Surveys*, 53(4), 2020.

J. Kim and W. Winkler. Multiplicative Noise for Masking Continuous Data. *Statistics*, 1:9, 2003.

S. J. Kim, S. Kang, Y. Choi, M. Choi, and M. Hong. Augmented-reality Survey: from Concept to Application. *KSII Transactions on Internet and Information Systems*, 11 (2):982–1004, 2017.

J. Konecny, H. B. McMahan, D. Ramage, and P. Richtárik. Federated Optimization: Distributed Machine Learning for On-Device Intelligence. *arXiv preprint arXiv:1610.02527*, 2016.

M. Köppen. The Curse of Dimensionality. In *Proc. Online World Conference on Soft Computing in Industrial Applications*, 2000.

R. Kumar, V. V. Phoha, and A. Serwadda. Continuous Authentication of Smartphone Users by Fusing Typing, Swiping and Phone Movement Patterns. In *Proc. IEEE International Conference on Biometrics Theory, Applications and Systems*, 2016.

R. D. Labati, V. Piuri, and F. Scotti. Biometric Privacy Protection: Guidelines and Technologies. In *Proc. International Conference on E-Business and Telecommunications*, 2011.

Z. Lai, Y. C. Hu, Y. Cui, L. Sun, and N. Dai. Furion: Engineering High-quality Immersive Virtual Reality on Today's Mobile Devices. In *Proc. Annual International Conference on Mobile Computing and Networking*, 2017.

P. Lamb, A. Millar, and R. Fuentes. Swipe Dynamics as a Means of Authentication: Results from a Bayesian Unsupervised Approach. In *Proc. IEEE International Joint Conference on Biometrics*, 2020.

B. Li, W. Cui, W. Wang, L. Zhang, Z. Chen, and M. Wu. Two-stream Convolution Augmented Transformer for Human Activity Recognition. In *Proc. AAAI Conference on Artificial Intelligence*, 2021a.

G. Li and P. Bours. Studying WiFi and Accelerometer Data based Authentication Method on Mobile Phones. In *Proc. International Conference on Biometric Engineering and Applications*, 2018.

N. Li and N. Ti. T-closeness: Privacy beyond K-anonymity and L-diversity. In *Proc. Conference on Data Engineering*, 2007.

Q. Li and G. Cao. Efficient and Privacy-preserving Data Aggregation in Mobile Sensing. In *Proc. IEEE International Conference on Network Protocols*, 2012.

S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y.-X. Wang, and X. Yan. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. *Advances in Neural Information Processing Systems*, 32, 2019.

Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 2021b.

Z. Lin, M. Hewett, and R. B. Altman. Using Binning to Maintain Confidentiality of Medical Data. In *Proc. AMIA Symposium*, 2002.

C. Liu, S. Chen, S. Zhou, J. Guan, and Y. Ma. A Novel Privacy Preserving Method for Data Publication. *Information Sciences*, 501:421–435, 2019.

S. K. Lo, Q. Lu, C. Wang, H.-Y. Paik, and L. Zhu. A Systematic Literature Review on Federated Machine Learning: From a Software Engineering Perspective. *ACM Computing Surveys*, 54(5), 2021.

L. Lu and Y. Liu. Safeguard: User Reauthentication on Smartphones via Behavioral Biometrics. *IEEE Transactions on Computational Social Systems*, 2(3):53–64, 2015.

D. G. Luca and M. Alberto. From Proximity to Accurate Indoor Localization for Context Awareness in Mobile Museum Guides. In *Proc. ACM International Conference on Mobile Human-Computer Interaction*, 2016.

L. Lyu, J. C. Bezdek, Y. W. Law, X. He, and M. Palaniswami. Privacy-preserving Collaborative Fuzzy Clustering. *Data & Knowledge Engineering*, 116:21–41, 2018.

Y. Ma, S. Arshad, S. Muniraju, E. Torkildson, E. Rantala, K. Doppler, and G. Zhou. Location-and Person-Independent Activity Recognition with WiFi, Deep Neural Networks and Reinforcement Learning. *ACM Transactions on Internet of Things*, 2(1): 1–25, 2021.

A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. L -diversity: Privacy beyond K -anonymity. *ACM Transactions on Knowledge Discovery from Data*, 1(1):3, 2007.

U. Mahbub, S. Sarkar, V. M. Patel, and R. Chellappa. Active User Authentication for Smartphones: A Challenge Data Set and Benchmark Results. In *Proc. IEEE International Conference on Biometrics Theory, Applications and Systems*, 2016.

E. Maiorana, P. Campisi, N. González-Carballo, and A. Neri. Keystroke Dynamics Authentication for Mobile Phones. In *Proc. ACM Symposium on Applied Computing*, 2011.

S. Majumder and M. J. Deen. Smartphone Sensors for Health Monitoring and Diagnosis. *Sensors*, 19(9), 2019.

M. Malekzadeh, R. G. Clegg, C. andrea, and H. Haddadi. Protecting Sensory Data against Sensitive Inferences. In *Proc. Workshop on Privacy by Design in Distributed Systems*, 2018.

D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar, et al. *Handbook of Fingerprint Recognition*, volume 2. Springer, 2009.

R. Mao, H. Ji, D. Cheng, X. Wang, Y. Wang, and D. Sun. Implicit Continuous Authentication Model Based on Mobile Terminal Touch Behavior. In *Proc. IEEE Symposium on Computers and Communications*, 2022.

M. D. Marsico and A. Mecca. A Survey on Gait Recognition via Wearable Sensors. *ACM Computing Surveys*, 52(4):1–39, 2019.

H. B. McMahan, D. Ramage, K. Talwar, and L. Zhang. Learning Differentially Private Recurrent Language Models. *arXiv preprint arXiv:1710.06963*, 2017.

L. R. Medsker and L. Jain. Recurrent Neural Networks. *Design and Applications*, 5: 64–67, 2001.

T. Meena and K. Sarawadekar. Gender Recognition using In-built Inertial Sensors of Smartphone. In *Proc. IEEE Region 10 Conference*, pages 462–467, 2020.

G. Melo, L. Oliveira, D. Schneider, and J. de Souza. Towards an Observatory for Mobile Participatory Sensing Applications. In *Proc. International Conference on Computer Supported Cooperative Work in Design*, 2017.

P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch. An Overview of Privacy-enhancing Technologies in Biometric Recognition. *arXiv preprint arXiv:2206.10465*, 2022a.

P. Melzi, R. Tolosana, and R. Vera-Rodriguez. ECG Biometric Recognition: Review, System Proposal and Benchmark Evaluation. *IEEE Access*, 2022b.

P. Melzi, H. O. Shahreza, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, J. Fierrez, S. Marcel, and C. Busch. Multi-IVE: Privacy Enhancement of Multiple Soft-Biometrics in Face Embeddings. In *Proc. Winter Conference on Applications of Computer Vision Workshops*, 2023.

W. Meng, W. Li, and D. S. Wong. Enhancing Touch Behavioral Authentication via Cost-based Intelligent Mechanism on Smartphones. *Multimedia Tools and Applications*, 77 (23):30167–30185, 2018a.

W. Meng, Y. Wang, D. S. Wong, S. Wen, and Y. Xiang. TouchWB: Touch Behavioral User Authentication based on Web Browsing on Smartphones. *Journal of Network and Computer Applications*, 117:1–9, 2018b.

S. Merugu and Joydeep Ghosh. Privacy-preserving Distributed Clustering using Generative Models. In *Proc. IEEE International Conference on Data Mining*, 2003.

O. Miguel-Hurtado, S. Stevenage, C. Bevan, and R. Guest. Predicting Sex as a Soft-biometrics from Device Interaction Swipe Festures. *Pattern Recognition Letters*, 79: 44–51, 2016.

S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang. Biometrics Recognition using Deep Learning: A Survey. *Artificial Intelligence Review*, pages 1–49, 2023.

V. Mirjalili, S. Raschka, A. Namboodiri, and A. Ross. Semi-Adversarial Networks: Convolutional Autoencoders for Imparting Privacy to Face Images. In *Proc. International Conference on Biometrics*, 2018.

V. Mirjalili, S. Raschka, and A. Ross. PrivacyNet: Semi-Adversarial Networks for Multi-attribute Face Privacy. *IEEE Transactions on Image Processing*, 29:9400–9412, 2020.

I. Mironov, O. Pandey, O. Reingold, and S. Vadhan. Computational Differential Privacy. In *Proc. International Cryptology Conference*, 2009.

K. Mivule. Utilizing Noise Addition for Data Privacy, an Overview. *arXiv preprint arXiv:1309.3958*, 2013.

S. Mondal and P. Bours. A Study on Continuous Authentication Using a Combination of Keystroke and Mouse Biometrics. *Neurocomputing*, 2017.

A. Morales, J. Fierrez, M. Gomez-Barrero, J. Ortega-Garcia, R. Daza, J. V. Monaco, J. Montalvão, J. Canuto, and A. George. KBOC: Keystroke Biometrics Ongoing Competition. In *Proc. International Conference on Biometrics Theory, Applications and Systems*, 2016.

A. Morales, J. Fierrez, R. Tolosana, J. Ortega-Garcia, J. Galbally, M. Gomez-Barrero, A. Anjos, and S. Marcel. Keystroke Biometrics Ongoing Competition. *IEEE Access*, 4:7736–7746, 2016.

A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana. SensitiveNets: Learning Agnostic Representations with Application to Face Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2158–2164, 2020.

C. Murphy, J. Huang, D. Hou, and S. Schuckers. Shared Dataset on Natural Human-Computer Interaction to Support Continuous Authentication Research. In *Proc. International Joint Conference on Biometrics*, 2017.

L. Na, C. Yang, C. Lo, F. Zhao, Y. Fukuoka, and A. Aswani. Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning. *JAMA Network Open*, 1(8):e186040–e186040, 2018.

K. Nandakumar and A. K. Jain. Biometric Template Protection: Bridging the Performance Gap between Theory and Practice. *IEEE Signal Processing Magazine*, 32(5): 88–100, 2015.

T. Neal and D. Woodard. A Gender-specific Behavioral Analysis of Mobile Device Usage Data. In *Proc. International Conference on Identity, Security and Behavior Analysis*, 2018.

M. E. Nergiz, M. Atzori, and C. Clifton. Hiding the Presence of Individuals from Shared Databases. In *Proc. ACM SIGMOD International Conference on Management of Data*, 2007.

T. T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi. The Largest Inertial Sensor-based Gait Database and Performance Evaluation of Gait-based Personal Authentication. *Pattern Recognition*, 47(1):228–237, 2014.

K.-T. Nguyen, T.-L. Vo-Tran, D.-T. Dinh, and M.-T. Tran. Gait Recognition with Multi-Region Size Convolutional Neural Network for Authentication with Wearable

Sensors. In *Proc. International Conference on Future Data and Security Engineering*, 2017.

T. Nguyen, A. Roy, and N. Memon. Kid on the Phone! Toward Automatic Detection of Children on Mobile Devices. *Computers & Security*, 84:334–348, 2019.

N. Niknejad, W. B. Ismail, A. Mardani, H. Liao, and I. Ghani. A Comprehensive Overview of Smart Wearables: The State of the Art Literature, Recent Advances and Future Challenges. *Engineering Applications of Artificial Intelligence*, 90:103529, 2020.

R. Nussbaum, C. Kelly, E. Quinby, A. Mac, B. Parmanto, and B. E. Dicianno. Systematic Review of Mobile Health Applications in Rehabilitation. *Archives of Physical Medicine and Rehabilitation*, 100(1):115–127, 2019.

L. O'Gorman. Comparing Passwords, Tokens and Biometrics for User Authentication. *IEEE*, 91(12):2021–2040, 2003.

S. R. M. Oliveira and O. R. Zaıane. Privacy Preserving Frequent Itemset Mining. In *Proc. IEEE International Conference on Privacy, Security and Data Mining*, 2002.

F. J. Ordóñez and D. Roggen. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors*, 16(1):115, 2016.

S. A. Osia, A. S. Shamsabadi, S. Sajadmanesh, A. Taheri, K. Katevas, H. R. Rabiee, N. D. Lane, and H. Haddadi. A Hybrid Deep Learning Architecture for Privacy-Preserving Mobile Analytics. *IEEE Internet of Things Journal*, 7(5):4505–4518, 2020.

K. Palin, A. M. Feit, S. Kim, P. O. Kristensson, and A. Oulasvirta. How Do People Type on Mobile Devices? Observations from a Study with 37,000 Volunteers. In *Proc. International Conference on Human-Computer Interaction with Mobile*, 2019.

I. Papavasileiou, Z. Qiao, C. Zhang, W. Zhang, J. Bi, and S. Han. GaitCode: Gait-based Continuous Authentication using Multimodal Learning and Wearable Sensors. *Smart Health*, 19:100162, 2021.

A. A. Parmar, U. P. Rao, and D. R. Patel. Blocking Based Approach for Classification Rule Hiding to Preserve the Privacy in Database. In *Proc. International Symposium on Computer Science and Society*, 2011.

V. M. Patel, N. K. Ratha, and R. Chellappa. Cancelable Biometrics: A Review. *IEEE Signal Processing Magazine*, 32(5):54–65, 2015.

V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello. Continuous User Authentication on Mobile Devices: Recent Progress and Remaining Challenges. *IEEE Signal Processing Magazine*, 33(4):49–61, 2016.

N. Phan, Y. Wang, X. Wu, and D. Dou. Differential Privacy Preservation for Deep Auto-Encoders: an Application of Human Behavior Prediction. In *Proc. AAAI Conference on Artificial Intelligence*, 2016.

J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng. A Survey of Machine Learning for Big Data Processing. *EURASIP Journal on Advances in Signal Processing*, 2016(1):1–16, 2016.

R. Ramachandra and C. Busch. Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey. *ACM Computing Surveys*, 50(1):1–37, 2017.

V. Rastogi, D. Suciu, and S. Hong. The Boundary Between Privacy and Utility in Data Publishing. In *Proc. International Conference on Very Large Data Bases*, 2007.

J. Ren, G. Wu, and L. Yao. A Sensitive Data Aggregation Scheme for Body Sensor Networks based on Data Hiding. *Personal and Ubiquitous Computing*, 17(7):1317–1329, 2013.

C. Riederer, S. Zimmeck, C. Phanord, A. Chaintreau, and S. Bellovin. I Don't Have a Photograph, but You Can Have my Footprints. Revealing the Demographics of Location Data. In *Proc. ACM on Conference on Online Social Networks*, 2015.

S. Romero-Tapiador, B. Lacruz-Pleguezuelos, R. Tolosana, G. Freixer, R. Daza, C. M. Fernández-Díaz, E. Aguilar-Aguilar, J. Fernández-Cabezas, S. Cruz Gil, S. Molina-Arranz, M. C. Crespo, T. Laguna-Lobo, L. J. Marcos-Zambrano, R. Vera-Rodriguez, J. Fierrez, A. Ramírez de Molina, J. Ortega-Garcia, I. Espinosa-Salinas, A. Morales, and E. Carrillo de Santa Pau. AI4FoodDB: A Database for Personalized e-Health Nutrition and Lifestyle through Wearable Devices and Artificial Intelligence. *Database*, 2023.

A. Ross and A. Jain. Biometric Sensor Interoperability: A Case Study in Fingerprints. In *Proc. International Workshop on Biometric Authentication*, 2004a.

A. Ross and A. K. Jain. Multimodal Biometrics: An Overview. In *Proc. European Signal Processing Conference*, 2004b.

A. Sabir, H. Maghdid, S. Asaad, M. Ahmed, and A. Asaad. Gait-based Gender Classification using Smartphone Accelerometer Sensor. In *Proc. International Conference on Frontiers of Signal Processing*, 2019.

D. Sadhya and S. K. Singh. Privacy Preservation for Soft Biometrics based Multimodal Recognition System. *Computers & Security*, 58:160–179, 2016.

D. Saha and A. Mukherjee. Pervasive Computing: a Paradigm for the 21st Century. *Computer*, 36(3):25–31, 2003.

M. Salehan and A. Negahban. Social Networking on Smartphones: When Mobile Phones Become Addictive. *Computers in Human Behavior*, 29(6):2632–2639, 2013.

K. Sampigethaya, L. Huang, M. Li, R. Poovendran, K. Matsuura, and K. Sezaki. CARAVAN: Providing Location Privacy for VANET. Technical report, Washington University Seattle Dept of Electrical Engineering, 2005.

D. Santani, T. Do, F. Labhart, S. Landolt, E. Kuntsche, and D. Gatica-Perez. DrinkSense: Characterizing Youth Drinking Behavior Using Smartphones. *IEEE Transactions on Mobile Computing*, 17(10):2279–2292, 2018.

M. Santopietro, R. Vera-Rodriguez, R. Guest, A. Morales, and A. Acien. Assessing the Quality of Swipe Interactions for Mobile Biometric Systems. In *Proc. IEEE International Joint Conference on Biometrics*, 2020.

P. Saravanan, S. Clarke, D. H. Chau, and H. Zha. Latentgesture: Active User Authentication through Background Touch Analysis. In *Proc. International Symposium of Chinese CHI*, 2014.

L. Scherrer, M. Tomko, P. Ranacher, and R. Weibel. Travelers or Locals? Identifying Meaningful Sub-populations from Human Movement Data in the Absence of Ground Truth. *EPJ Data Science*, 7:1–21, 2018.

F. Schroff, D. Kalenichenko, and J. Philbin. FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

A. Sepas-Moghaddam and A. Etemad. Deep Gait Recognition: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

S. Servia-Rodríguez, K. Rachuri, C. Mascolo, P. Rentfrow, N. Lathia, and G. Sandstrom. Mobile Sensing at the Service of Mental Well-Being: A Large-Scale Longitudinal Study. In *Proc. International Conference on World Wide Web*, 2017.

A. Serwadda, V. V. Phoha, and Z. Wang. Which Verifiers Work?: A Benchmark Evaluation of Touch-based Authentication Algorithms. In *Proc. IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2013.

C. E. Shannon. A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

V. Sharma and R. Enbody. User Authentication and Identification from User Interface Interactions on Touch-enabled Devices. In *Proc. ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2017.

C. Shen, Y. Zhang, X. Guan, and R. A. Maxion. Performance Analysis of Touch-interaction Behavior for Active Smartphone Authentication. *IEEE Transactions on Information Forensics and Security*, 11(3):498–513, 2015.

C. Shen, Y. Li, Y. Chen, X. Guan, and R. A. Maxion. Performance analysis of multi-motion sensor behavior for active smartphone authentication. *IEEE Transactions on Information Forensics and Security*, 13(1):48–62, 2018.

R. Shokri, G. Theodorakopoulos, J. L. Boudec, and J. Hubaux. Quantifying Location Privacy. In *Proc. IEEE Symposium on Security and Privacy*, 2011.

P. Siirtola, J. Komulainen, and V. Kellokumpu. Effect of Context in Swipe Gesture-based Continuous Authentication on Smartphones. In *Proc. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2018.

S. Singh, D. M. Shila, and G. Kaiser. Side Channel Attack on Smartphone Sensors to Infer Gender of the User: Poster Abstract. In *Proc. Conference on Embedded Networked Sensor Systems*, 2019.

Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani. HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone

Users. *IEEE Transactions on Information Forensics and Security*, 11(5):877–892, 2015.

S. Sprager and M. B. Juric. Inertial Sensor-based Gait Recognition: A Review. *Sensors*, 15(9):1–39, 2015.

G. Stragapede, R. Vera-Rodriguez, R. Tolosana, and A. Morales. BehavePassDB: Benchmarking Mobile Behavioral Biometrics. *Pattern Recognition*, 2022a.

G. Stragapede, R. Vera-Rodriguez, R. Tolosana, A. Morales, A. Acien, and G. Le Lan. Mobile Behavioral Biometrics for Passive Authentication. *Pattern Recognition Letters*, 2022b.

G. Stragapede, R. Vera-Rodriguez, R. Tolosana, A. Morales, J. Fierrez, J. Ortega-Garcia, S. Rasnayaka, S. Seneviratne, V. Dissanayake, J. Liebers, et al. IJCB 2022 Mobile Behavioral Biometrics Competition (MobileB2C). In *Proc. International Joint Conference on Biometrics (IJCB)*, pages 1–7. IEEE, 2022c.

G. Stragapede, P. Delgado-Santos, R. Tolosana, R. Vera-Rodriguez, R. Guest, and A. Morales. Mobile Keystroke Biometrics Using Transformers. In *Proc. International Conference on Automatic Face and Gesture Recognition*, 2023a.

G. Stragapede, P. Delgado-Santos, R. Tolosana, R. Vera-Rodriguez, R. Guest, and A. Morales. TypeFormer: Transformers for Mobile Keystroke Biometrics. *Under Review in ACM Transactions Computer-Human Interaction*, 2023b.

R. Subramanian and S. Sarkar. Evaluation of Algorithms for Orientation Invariant Inertial Gait Matching. *IEEE Transactions on Information Forensics and Security*, 14(2):304–318, 2018.

L. Sun, D. Zhang, B. Li, B. Guo, and S. Li. Activity Recognition on an Accelerometer

Embedded Mobile Phone with Varying Positions and Orientations. *Ubiquitous Intelligence and Computing*, 6406:548–562, 2010.

Y. Sun, H. Ceker, and S. Upadhyaya. Shared Keystroke Dataset for Continuous Authentication. In *Proc. International Workshop on Information Forensics and Security*, 2016.

Y. Sun, J. Tang, X. Shu, Z. Sun, and M. Tistarelli. Facial Age Synthesis with Label Distribution-Guided Generative Adversarial Network. *IEEE Transactions on Information Forensics and Security*, 15:2679–2691, 2020.

D. Svantesson and R. Clarke. Privacy and Consumer Risks in Cloud Computing. *Computer Law & Security Review*, 26(4):391–397, 2010.

J. Svoboda, J. Masci, and M. Bronstein. Palmprint Recognition Via Discriminative Index Learning. In *Proc. International Conference on Pattern Recognition*, 2016.

L. Sweeney. K-anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

Z. Syed, J. Helmick, S. Banerjee, and B. Cukic. 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, 2019.

Y. Tay, M. Dehghani, D. Bahri, and D. Metzler. Efficient Transformers: A Survey. *ACM Computing Surveys*, 55(6):1–28, 2022.

I. Tayfur and M. A. Afacan. Reliability of Smartphone Measurements of Vital Parameters: A Prospective Study using a Reference Method. *The American Journal of Emergency Medicine*, 37(8):1527–1530, 2019.

M. Templ, A. Kowarik, and B. Meindl. Statistical Disclosure Control for Micro-Data

Using the R Package sdcMicro. *Journal of Statistical Software, Articles*, 67(4):1–36, 2015.

P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper. Suppressing Gender and Age in Face Templates using Incremental Variable Elimination. In *Proc. International Conference on Biometrics*, 2019.

E. Thomaz, I. Essa, and G. D. Abowd. A Practical Approach for Recognizing Eating Moments with Wrist-mounted Inertial Sensing. In *Proc. ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015.

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Exploring Recurrent Neural Networks for On-line Handwritten Signature Biometrics. *IEEE Access*, 6:5128–5138, 2018.

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Reducing the Template Aging Effect in On-Line Signature Biometrics. *IET Biometrics*, 8(6):422–430, June 2019.

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia. Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection. *Information Fusion*, 64:131–148, 2020.

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. BioTouchPass2: Touchscreen Password Biometrics Using Time-Aligned Recurrent Neural Networks. *IEEE Transactions on Information Forensics and Security*, 15:2616–2628, 2020.

R. Tolosana, R. Vera-Rodriguez, R. Guest, J. Fierrez, and J. Ortega-Garcia. Exploiting Complexity in Pen-and Touch-based Signature Biometrics. *International Journal on Document Analysis and Recognition*, 23(2):129–141, 2020.

R. Tolosana, P. Delgado-Santos, P.-U. andres, R. Vera-Rodriguez, J. Fierrez, and A. Morales. DeepWriteSYN: On-Line Handwriting Synthesis via Deep Short-Term Representations. In *Proc. AAAI Conference on Artificial Intelligence*, 2021a.

R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. DeepSign: Deep On-Line Signature Verification. *IEEE Transactions on Biometrics, Behavior and Identity Science*, 3(2):229–239, 2021b.

R. Tolosana, J. C. Ruiz-Garcia, R. Vera-Rodriguez, J. Herreros-Rodriguez, S. Romero-Tapiador, A. Morales, and J. Fierrez. Child-Computer Interaction with Mobile Devices: Recent Works, New Dataset and Age Detection. *IEEE Transactions on Emerging Topics in Computing*, 10(4):2042–2054, 2022a.

R. Tolosana, R. Vera-Rodriguez, C. Gonzalez-Garcia, J. Fierrez, A. Morales, J. Ortega-Garcia, J. C. Ruiz-Garcia, S. Romero-Tapiador, S. Rengifo, M. Caruana, et al. SVC-onGoing: Signature Verification Competition. *Pattern Recognition*, 127:1–14, 2022b.

F. Tramèr and D. Boneh. Differentially Private Learning Needs Better Features (or Much More Data). *arXiv preprint arXiv:2011.11660*, 2020.

L. Tran and D. Choi. Data Augmentation for Inertial Sensor-based Gait Deep Neural Network. *IEEE Access*, 8:12364–12378, 2020.

L. Tran, T. Hoang, T. Nguyen, H. Kim, and D. Choi. Multi-Model Long Short-Term Memory Network for Gait Recognition using Window-based Data Segment. *IEEE Access*, 9:23826–23839, 2021.

M. Trauring. Automatic Comparison of Finger-ridge Patterns. *Nature*, 197:938–940, 1963.

S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou. A

Hybrid Approach to Privacy-Preserving Federated Learning. In *Proc. ACM Workshop on Artificial Intelligence and Security*, 2019.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention Is All You Need. In *Proc. Advances in Neural Information Processing Systems*, 2017.

G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Pediaditis, and M. Tsiknakis. The Mobiact Dataset: Recognition of Activities of Daily Living using Smartphones. In *Proc. International Conference on Information and Communication Technologies for Ageing Well and e-Health*, 2016.

V. Verykios, E. Bertino, I. Fovino, L. Provenza, Y. Saygin, and Y. Theodoridis. State-of-the-Art in Privacy Preserving Data Mining. *SIGMOD Rec.*, 33(1):50–57, 2004.

M. Villani, C. Tappert, G. Ngo, J. Simone, H. S. Fort, and S.-H. Cha. Keystroke Biometric Recognition Studies on Long-Text Input Under Ideal and Application-Oriented Conditions. In *Proc. Conference on Computer Vision and Pattern Recognition Workshop*, 2006.

I. Wagner and D. Eckhoff. Technical Privacy Metrics: a Systematic Survey. *ACM Computing Surveys (CSUR)*, 51(3):1–38, 2018.

C. Wan, L. Wang, and V. V. Phoha. A Survey on Gait Recognition. *ACM Computing Surveys*, 51(5), Aug. 2018.

N. Wan and G. Lin. Classifying Human Activity Patterns from Smartphone Collected GPS Data: A Fuzzy Classification and Aggregation Approach. *Transactions in GIS*, 20(6):869–886, 2016.

L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette Analysis-Based Gait Recognition for

Human Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003.

X. Wang, T. Yu, O. Mengshoel, and P. Tague. Towards Continuous and Passive Authentication across Mobile Devices: an Empirical Study. In *Proc. ACM Conference on Security and Privacy in Wireless and Mobile Networks*, 2017.

X. Wang, Y. Zhao, and F. Pourpanah. Recent Advances in Deep Learning. *International Journal of Machine Learning and Cybernetics*, 11:747–750, 2020.

Y. Watanabe and M. Kimura. Gait Identification and Authentication using LSTM based on 3-axis Accelerations of Smartphone. *Procedia Computer Science*, 176:3873–3880, 2020.

Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun. Transformers in Time Series: A Survey. *arXiv preprint arXiv:2202.07125)*, 2022.

J. Wieringa, P. Kannan, X. Ma, T. Reutterer, H. Risselada, and B. Skiera. Data Analytics in a Privacy-concerned World. *Journal of Business Research*, 122:915–925, 2021.

R. C. Wong, J. Li, A. W. Fu, and K. Wang. ($\alpha$, k)-Anonymity: an Enhanced k-Anonymity Model for Privacy Preserving Data Publishing. In *Proc. ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006.

R. Wright and L. Keith. Wearable Technology: If the Tech fits, Wear Tt. *Journal of Electronic Resources in Medical Libraries*, 11(4):204–216, 2014.

H. Wu, J. Xu, J. Wang, and M. Long. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In *Proc. Advances in Neural Information Processing Systems*, 2021.

L. Wu, L. Yang, Z. Huang, Y. Wang, Y. Chai, X. Peng, and Y. Liu. Inferring Demographics from Human Trajectories and Geographical Context. *Computers, Environment and Urban Systems*, 77:101368, 2019.

X. Xiao and Y. Tao. M-invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets. In *Proc. ACM SIGMOD International Conference on Management of Data*, 2007.

H. Xu, Y. Zhou, and M. R. Lyu. Towards Continuous and Passive Authentication via Touch Biometrics: An Experimental Study on Smartphones. In *Proc. Symposium On Usable Privacy and Security*, 2014.

M. Yang, T. Zhu, Y. Xiang, and W. Zhou. Density-Based Location Preservation for Mobile Crowdsensing With Differential Privacy. *IEEE Access*, 6:14779–14789, 2018.

Q. Yang, Y. Liu, T. Chen, and Y. Tong. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10 (2):1–19, 2019.

Y. Yuan, M. Raubal, and Y. Liu. Correlating Mobile Phone Usage and Travel Behavior–A Case Study of Harbin, China. *Computers, Environment and Urban Systems*, 36(2): 118–130, 2012.

V. Zaliva, W. Melicher, S. Saha, and J. Zhang. Passive User Identification using Sequential Analysis of Proximity Information in Touchscreen Usage Patterns. In *Proc. IEEE International Conference on Mobile Computing and Ubiquitous Networking*, 2015.

D. Zhang, L. Yao, K. Chen, Z. Yang, X. Gao, and Y. Liu. Preventing Sensitive Information Leakage from Mobile Sensor Signals via Integrative Transformation. *IEEE Transactions on Mobile Computing*, 21(12):4517–4528, 2021.

H. Zhang, V. M. Patel, M. Fathy, and R. Chellappa. Touch Gesture-based Active User Authentication using Dictionaries. In *Proc. IEEE Winter Conference on Applications of Computer Vision*, 2015.

N. Zhang, J. Wang, Z. Hong, C. Zhao, X. Qu, and J. Xiao. DT-SV: A Transformer-based Time-domain Approach for Speaker Verification. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2022a.

Q. Zhang, N. Koudas, D. Srivastava, and T. Yu. Aggregate Query Answering on Anonymized Tables. In *Proc. International Conference on Data Engineering*, 2007.

X. Zhang, X. Jin, K. Gopalswamy, G. Gupta, Y. Park, X. Shi, H. Wang, D. C. Maddix, and Y. Wang. First De-Trend then Attend: Rethinking Attention for Time-Series Forecasting. In *NeurIPS 2022 Workshop on All Things Attention: Bridging Different Perspectives on Attention*, 2022b.

Y. Zhong and Y. Deng. Sensor Orientation Invariant Mobile Gait Biometrics. In *Proc. IEEE International Joint Conference on Biometrics*, 2014.

H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In *Proc. AAAI Conference on Artificial Intelligence*, 2021.

Q. Zou, Y. Wang, Q. Wang, Y. Zhao, and Q. Li. Deep Learning-based Gait Recognition using Smartphones in the Wild. *IEEE Transactions on Information Forensics and Security*, 15:3197–3212, 2020.