N°d'ordre NNT : xxx

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
(Institut National des Sciences Appliquées, INSA - Lyon)

Ecole Doctorale N° 205
(Interdisciplinaire Sciences Santé)

Spécialité/ discipline de doctorat :
Ingénierie biomédicale, biotechnologie

Soutenue publiquement le 28/10/2021, par:
Théo Jourdan

Privacy and transparency in learning systems for healthcare

Devant le jury composé de :

Fossati, Caroline	Professeure des Universités, Institut Fresnel	Rapporteure
Vincent, Emmanuel	Directeur de Recherche, INRIA Nancy	Rapporteur
Bellet, Aurélien	Chargé de Recherche, INRIA Lille	Examinateur
Ben Mokhtar, Sonia	Directrice de Recherche, LIRIS	Examinatrice
Dieterlen, Alain	Professeur des Universités, IRIMAS	Examinateur
Frindel, Carole	Maître de Conférences, INSA Lyon	Co-directrice de thèse
Boutet, Antoine	Maître de Conférences, INSA Lyon	Co-directeur de thèse

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **https://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage secretariat@edchimie-lyon.fr	**M. Stéphane DANIELE** C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne **directeur@edchimie-lyon.fr**
E.E.A.	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **https://edeea.universite-lyon.fr** Sec. : Stéphanie CAUVIN Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	**M. Philippe DELACHARTRE** INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 **philippe.delachartre@insa-lyon.fr**
E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	**M. Philippe NORMAND** Université Claude Bernard Lyon 1 UMR 5557 Lab. d'Ecologie Microbienne Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69 622 Villeurbanne CEDEX **philippe.normand@univ-lyon1.fr**
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://ediss.universite-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	**Mme Sylvie RICARD-BLUM** Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (ICBMS) - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 **sylvie.ricard-blum@univ-lyon1.fr**
INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	**M. Hamamache KHEDDOUCI** Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 **hamamache.kheddouci@univ-lyon1.fr**
Matériaux	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 yann.de-ordenana@ec-lyon.fr	**M. Stéphane BENAYOUN** Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 **stephane.benayoun@ec-lyon.fr**
MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon mega@insa-lyon.fr	**M. Jocelyn BONJOUR** INSA Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69621 Villeurbanne CEDEX **jocelyn.bonjour@insa-lyon.fr**
ScSo	**ScSo*** **https://edsciencessociales.universite-lyon.fr** Sec. : Mélina FAVETON INSA : J.Y. TOUSSAINT Tél : 04.78.69.77.79 melina.faveton@univ-lyon2.fr	**M. Christian MONTES** Université Lumière Lyon 2 86 Rue Pasteur 69365 Lyon CEDEX 07 **christian.montes@univ-lyon2.fr**

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Résumé

Avec le développement de l'Internet des objets (IdO), les smartphones et les capteurs sont désormais capables de fournir des informations sur l'activité de l'utilisateur et même sur sa physiologie. Cela a donc suscité un intérêt croissant de la part de la communauté scientifique, notamment dans le domaine de la e-santé avec des applications dans le suivi des patients en cours de rééducation pour offrir un suivi plus personnalisé. Cependant, outre le fait de guider le processus de rééducation, la production et la transmission de données IdO sont également exposées à des atteintes à la vie privée. En effet, la chaîne de traitement complexe de l'application IdO dans les soins de santé multiplie les risques de menaces sur la vie privée tout au long du cycle de vie des données IdO, comprenant la collecte, la transmission et le stockage, par un adversaire qui peut récupérer les données et ré-identifier ou révéler des informations sensibles des patients. Cette thèse s'articule autour des questions suivantes: Les données collectées sont-elles suffisamment protégées pour que personne ne puisse en abuser pour ré-identifier le propriétaire ou déduire des informations sensibles ? Les données protégées sont-elles encore suffisamment précises pour les applications de soins de santé telles que la rééducation ? Atteindre cet équilibre entre l'utilité des données et la protection de la vie privée est un défi important que nous étudions dans cette thèse sous différents angles. Plus précisément, la première partie se concentre sur le problème de l'anonymisation des données par le biais de la minimisation, tandis que la deuxième partie se concentre sur la prévention de l'inférence d'attributs sensibles par le biais d'une approche basée sur les Réseaux Génératifs Adversariaux pour assainir les données des capteurs et une approche exploitant les couches privées dans l'apprentissage fédéré.

Mots clés

traitement du signal, reconnaissance d'activité humaine, prédiction, apprentissage automatique, deep learning, apprentissage fédéré, réseaux génératifs adversariaux

Abstract

With the development of the Internet of Things (IoT), smartphones and sensors are now able to provide information about the user's activity and even their physiology. This has led to a growing interest from the scientific community, particularly in the field of e-health, with applications in the monitoring of patients undergoing rehabilitation in order to offer more personalised follow-up. However, in addition to guiding the rehabilitation process, the generation and transmission of IoT data is also vulnerable to privacy breaches. Indeed, the complex processing chain of the IoT application in healthcare multiplies the risk of privacy threats throughout the life cycle of IoT data, including collection, transmission and storage, by an adversary who can retrieve the data and re-identify or reveal sensitive patient information. This thesis focuses on the following questions: Is the data collected sufficiently protected so that no one can misuse it to re-identify the owner or infer sensitive information? Is the protected data still accurate enough for healthcare applications such as rehabilitation? Achieving balance between data utility and privacy protection is an important challenge that we explore in this thesis from different angles. More specifically, the first part focuses on the problem of data anonymisation through minimisation, while the second part focuses on preventing the inference of sensitive attributes through a Generative Adversarial Networks (GAN) to sanitise sensor data and an approach exploiting private layers in Federated Learning (FL).

Keywords

signal processing, human activity recognition, privacy, prediction, machine learning, deep learning, federated learning, generative adversarial networks

Publications

Journals

- Jourdan, T., Boutet, A., Bahi, A., Frindel, C., **Privacy-preserving IoT Framework for Activity Recognition in Personal Healthcare Monitoring**, 2020, *ACM Transactions on Computing for Healthcare* (see chapter III.1)
- Jourdan, T., Debs, N., Frindel, C., **The Contribution of Machine Learning in the Validation of Commercial Wearable Sensors for Gait Monitoring in Patients: A Systematic Review**, 2021, *Sensors* (see chapter II.2)
- Jourdan, T., Boutet, A., Frindel, C., **Vers la protection de la vie privée dans les objets connectés pour la reconnaissance d'activité en santé**, 2019, Revue des Sciences et Technologies de l'Information-Série TSI: Technique et Science Informatiques (see chapter III.1)

Conferences

- Jourdan, T., Boutet, A., Frindel, C., **Toward privacy in IoT mobile devices for activity recognition**, 2018, *ACM/EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)* (see chapter III.1)
- Boutet, A., Frindel, C., Gambs, S., Jourdan, T., Ngueveu, R.C., **DySan: Dynamically Sanitizing Motion Sensor Data Against Sensitive Inferences through Adversarial Networks**, 2021, *ACM Asia Conference on Computer and Communications Security (ASIA CCS)* (see chapter IV.1)
- Debs, N., Jourdan, T., Moukadem, A., Boutet, A., Frindel, C., **Motion sensor data anonymization by time-frequency filtering**, 2020, *European Signal Processing Conference (EUSIPCO)* (see chapter III.2)

Workshops

- Jourdan, T., Boutet, A., Frindel, C., **Toward privacy in IoT mobile devices for activity recognition**, 2018, *Privacy Preserving Machine Learning NeurIPS Workshop (PPML)*
- Boutet, A., Frindel, C., Gambs, S., Jourdan, T., Ngueveu, R.C., **Protecting motion sensor data against sensitive inferences**, 2019, *l'Atelier sur la Protection de la Vie Privée (APVP)*

- Jourdan, T., Boutet, A., Frindel, C., **Privacy Assessment of Federated Learning using Private Personalized Layers**, 2021, *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)* (see chapter IV.2)

Teaching symposium

- Debs, N., Peignier, S., Douarre, C., Jourdan, T., Rigotti, C., Frindel, C., **Apprendre l'apprentissage automatique : un retour d'expérience**, 2020, *Colloque de l'Enseignement des Technologies et des Sciences de l'Information et des Systèmes (CETSIS)*

Acknowledgements

Je tiens à remercier toutes les personnes qui, de près ou de loin, m'ont aidé à mener à bien ma thèse. Plutôt que dresser une liste exhaustive, j'addresserais dans cette page quelques mots à différents groupes de personnes sans forcément les nommer explicitement.

Je souhaite tout d'abord remercier mes directeurs de thèse, Carole Frindel et Antoine Boutet, pour leur soutien essentiel durant ces trois années, leur patience et leur pédagogie. Je leur suis extrêmement reconnaissant de m'avoir soutenue toujours avec bienveillance et de m'avoir toujours donné le temps nécessaire pour transmettre leur savoir et leur expertise.

J'aimerais ensuite remercier Caroline Fossati et Emmanuel Vincent qui m'ont fait l'honneur d'accepter de lire et juger mes travaux en qualité de rapporteures. J'aimerais aussi remercier vivement Aurélien Bellet, Sonia Ben Mokhtar et Alain Dieterlen d'avoir accepté d'évaluer mes travaux en tant qu'examinateurs.

Je remercie les membres du laboratoire Creatis, et particulièrement mes camarades et amis doctorants qui m'ont accompagné pendant ces années et contribué à la bonne ambiance quotidienne.

Je remercie les membres de l'équipe Privactics et les personnes de l'antenne d'Inria, avec qui j'ai partagé de nombreux repas et discussions passionnantes autour du meilleur café du campus.

Je tiens à remercier particulièrement mon colocataire et surtout ami qui sans le savoir a été essentiel pendant ces trois années au quotidien, je le remercie de m'avoir fait grandir de par le regard qu'il porte sur le monde.

Je souhaite également remercier mes amis insaliens et amis lyonnais pour leur soutien précieux, avec qui j'ai partagé beaucoup de mes plus beaux moments pendant ces trois années.

Je souhaite évidemment remercier mes amis du Chateau Double, chers à mon coeur, qui pendant ces trois années et depuis longtemps déjà m'ont aidé à devenir la personne que je suis.

Je remercie enfin profondément ma famille pour leur soutien et amour indéfectible, essentiel à mon équilibre.

Contents

List of Figures

List of Tables

List of Acronyms

ADLs Activities of Daily Living

BMI Body Mass Index

CNIL Commission Nationale de l'Informatique et des Libertés
CNN Convolutional Neural Network

FL Federated Learning

GAN Generative Adversarial Networks
GDPR General Data Protection Regulation

IMU Inertial Measurement Units
IoT Internet of Things

k-NN k-Nearest Neighbors

LDP Local Differential Privacy
LSTM Long Short-Term Memory

ML Machine Learning

ReLU Rectified Linear Unit
RF Random Forest

SGD Stochastic Gradient Descent
STFT Short-Time Fourier Transform
SVM Support Vector Machine

UWB Ultra-wideband

Chapter I

Introduction

Health is a field that is evolving at the pace of technological progress and where the penetration of digital technology is bringing unprecedented changes in usage and behaviour. A particularly important innovation is the emergence of the Internet of Things (IoT) based tools that can be connected sensors (synchronised with a mobile app) or mobile apps (which use smartphone sensors) to measure a number of lifestyle-related constants. Health domain and particularly physical monitoring is now being impacted by the development of this technology whether in terms of health benefits perspectives or the numerous privacy issues it generates. To introduce this topic, we first present the context of this thesis in Section I.1 based on the health benefits brought by IoT (*i.e.*, utility) and the increasing concerns about leakage of personal data (*i.e.*, privacy). In Section I.2, we describe the contributions of this thesis to address the utility and privacy trade-off presented previously in the same section.

I.1 Motivations

I.1.1 Toward personalized medecine

I.1.1.1 Clinical context

Gait activities such as walking or running are the most essential physical activities for a human, but lots of neurological diseases cause gait impairments. For example ischaemic stroke is a serious condition caused by an abrupt cessation of blood flow to a part of the brain. When blood flow stops, brain cells in the affected area of the brain die, depriving them of a vital supply of oxygen and nutrients. Stroke results in a neurological deficit that may affect motor skills and gait activities. Stroke is the second leading cause of death and third leading cause of disability, with one stroke occurring every 5 seconds worldwide, resulting in nearly 15 million cases of stroke each year [155]. The life-threatening and functional risks associated with stroke make it a global public health priority. More widely, the ageing of the population in most developed countries leads to an increase of chronic diseases and becomes a social problem. Leading to hospitalization and long-term disability, it consumes a large amount of healthcare resources. Chronic diseases are the health conditions among the most costly in the United States with almost half of the American population that suffer from one of them [251]. With this percentage increasing over the years, hospitals and specialized health centers become crowded with patients and the time of hospitalization is not compatible with a long rehabilitation which is generally made at home [344]. In addition to recovering from neurological diseases, physical rehabilitation is also necessary for the elderly to remain active in the late ages.

Usually the rehabilitation is only made with human resources and can lead to low precision in the follow-up. Indeed, the clinician does not always have access to a complete and accurate feedback which leads to a lack of gait metrics that could help them to adapt the rehabilitation to the patient. Therefore, the effectiveness of current motor rehabilitation therapies can be questioned. For example concerning rehabilitation after a stroke, during the acute and sub-acute stages (<6 months after a stroke), patients receive rehabilitation

therapies in specialised health centres, consisting of a set of exercises with the aim of gaining maximum independence. After leaving the rehabilitation centre (*i.e.*, after entering the chronic stage, about 6 months after a stroke), only occasional medical appointments are scheduled to monitor the patient's progress. So during this chronic phase, the lack of continuous follow-up hinders optimal recovery: after returning home, about 65% of patients are unable to integrate the affected part of their body into activities of daily living again [58]. This points to the need for new therapeutic options that allow patients to train intensively and extensively after leaving the specialist centre, while ensuring the quality, effectiveness and safety of the therapy.

Different technological solutions emerged in order to overcome the lack of quantification and objectivity, such as motion capture with a camera or electronic walkway [297]. These technologies allow the clinician to have a quantitative monitoring of the rehabilitation with the acquisition of more indicators. As the data can be stored, the clinician can also follow the evolution of the movement of the patient through the different sessions. However, these kinds of technologies remain costly. The rehabilitation sessions need to be done in a dedicated laboratory, which requires patients to come. So there is no long time monitoring, recordings are generally done on a short distance which implies that gait patterns may be different from daily life [222]. These methods also requires computing skills and calibration. For these reasons, they are rather used for research rather than clinical monitoring [297].

I.1.1.2 Wearable sensors for rehabilitation

With the emergence of the IoT, more and more people are equipped with smartphones (60% of the French population) and other connected objects (activity monitoring bracelets, smartwatches) [309] that provide information on a person's activity and even on their physiology. Recent studies tend to show that the measurements of these sensors are sufficiently precise to return reliable information about the gait's user [89]. Wearable sensors consist of three main components: 1) hardware to detect and collect physiological or motion data, 2) communication hardware and software to relay the data to a remote center, and 3) data analysis techniques to extract clinically relevant information from the physiological and motion data [241]. Recently developed wearable systems incorporate sensors that use wireless, low-energy, low-cost, and high data rate communication technology. Common technologies include ZigBee, Bluetooth, Wi-Fi, and Ultra-wideband (UWB) pulse radio [348]. In healthcare, these wearable sensors are worn close to and/or on the surface of the skin, where they detect, analyze and transmit information about body signals (physiological signal, activity signal). In some cases, they provide immediate feedback (biofeedback) to the user [82, 83, 237]. These sensors exchange data with other users or connected devices, without human intervention and via wireless networks. This technology is now integrated in remote personal monitoring systems which is a complex framework with numerous elements interacting and collaborating with each other for a common task. A conceptual representation of a basic remote monitoring system is shown in Figure I.1.

Wearable sensors collect physiological and/or motion data to monitor the patient's condition ❶. When the device used is not directly included in the smartphone, wireless communication is used to transmit the patient's data to a mobile phone or access point ❷ and relay the information to a cloud (*i.e.*, a remote server potentially hosted by an untrusted third party) ❸ via the Internet. As a huge amount of data is generated per second, cloud computing helps in the storage and analysis of such data. Because the storage and calculations are performed on the cloud, the device is easy to use, accessible and inexpensive [45, 56]. Once collected, the patient data is processed to extract relevant clinical variables via signal processing, pattern recognition, data mining and other artificial intelligence-based methods

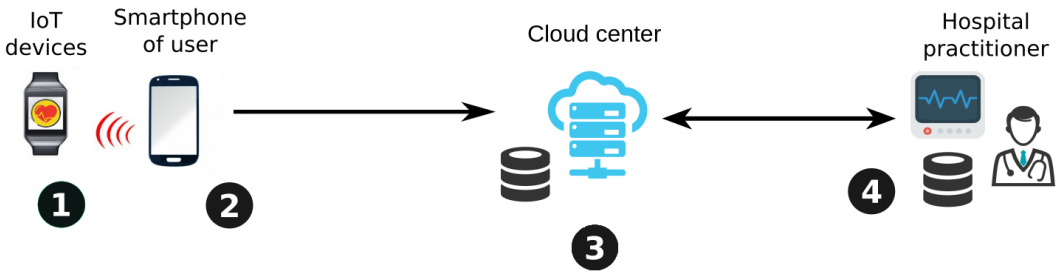


FIGURE I.1 – Illustration of a remote health monitoring system based on wearable sensors

[132]. In case of emergency (e.g., detected falls) an alarm message can be sent to an emergency service center ❹ to provide immediate assistance to patients. Family members and caregivers are alerted in case of an emergency but can also communicate regularly with the patient while remotely monitoring their clinical condition.

These technologies have therefore been the subject of much research in recent years in the medical field, particularly in chronic disease rehabilitation, with the aim of integrating them into the home activity monitoring process. Used consistently, wearable sensors can reduce assessment times and provide objective and quantifiable data on the physical and physiological capabilities of patients, complementing the expert judgement of rehabilitation health specialists. Sensors also have the potential to provide continuous information about activities of daily living. This wealth of information can then be used to fine-tune the patient's medical record, which can then lead to more targeted specific care (towards more personalised medicine). This potentially opens a new era for the field of signal processing applied to the study of chronic disease. In this context, many challenges remain for the use of wearable sensors in clinical routine, including the fact that current on-board sensors are not necessarily recognized as medical devices or that their clinical use is complex (need to build up patient cohorts outside hospitals).

Among the wearable sensors, the most commonly used are motion sensors (such as gyroscopes, accelerometers, pressure sensors, magnetometers) and Inertial Measurement Units (IMU), that are very useful for monitoring the activity of patients over the long term.

I.1.1.3 Motion sensors for different applications

Considerable research efforts have been made in recent years to assess the accuracy of these motion sensors in classifying Activities of Daily Living (ADLs) for home monitoring of the elderly and people with chronic diseases [105]. Studies have already shown that ADLs can be correctly identified in the elderly, either with accelerometers [203], or with pressure sensors in the shoe [274]. Fall detection devices have also been proposed, in the form of wearable sensors [116, 228], or sensors embedded in the telephone [4, 7]. Ongoing research is focused on the prevention of fall-related injuries. Motion devices have also been proposed to monitor imbalances in the gait of Parkinson's patients [31] and patients with dementia [289], stroke [331, 338], or to monitor potential seizures in epilepsy patients [318].

Clinical studies have shown that the use of motion sensors in home rehabilitation therapy is motivating by providing playful interfaces and encouraging patients to continue exercising regularly, which is beneficial in the long term [12, 249, 320]. Few studies already proposed contributions for home rehabilitation therapy. For example, in Parkinson's disease, it is essential to identify motor dysfunctions early enough to determine the optimal drug

dosage. The use of a sensor-based system therefore appears to be a promising approach to improve the clinical management of patients . Kostikis and al. [174] proposed a practical smartphone-based tool to accurately assess upper limb tremor, effective to remotely evaluate the patient's condition and communicate the results to the clinician. Lipsmeier and al. [191] used smartphone's sensors to monitor PD patients during 6 months and used digital biomarkers such as time spent walking or sit-to-stand transitions to assess disease status and treatment effects. Ferreira and al. [97] developed stroke rehabilitation exercises in a game-like structure using a smartphone in order to promote and evaluate different movements of the upper limbs. After a stroke, intensive rehabilitation is very important to recover motor functions. This motor activity monitoring is used as a feedback tool to guide the rehabilitation process, but also to collect clinically relevant data for rehabilitation staff regarding the patient's motor status, assess what the patient is able to endure, what type of exercise is beneficial to him and finally increase his motivation.

However, still most of the studies have been conducted in the laboratory, with a limited sample of patients. It would be necessary to validate more of these monitoring methods at home, under less controlled experimental conditions, and on a larger representative cohort.

I.1.1.4 Validation of the wearable sensors

In spite of their rapid development, the use of wearable sensors for the general public is sometimes contested in the medical community: the objections concern the quality of the data collected as well as the reliability of the technologies in a clinical context where the pathologies are very varied and even combined [43]. It still seems necessary to validate the sensors in order to remove any reluctance of the medical staff and to integrate them into a clinical setting dedicated to patient rehabilitation. A first step would be to define precise validation protocols - in consultation with the medical profession - adapted to the study of chronic pathologies. Indeed, many studies are content to validate sensors for a given medical application without even having tested them outside the laboratory [9, 33, 212], on a very limited number of patients [91, 185, 256], and over a relatively short time window (of the order of a few hours and even few minutes) [5, 220, 303]. In section II.2, we summarize a state-of-the-art between 2010 and 2020 in terms of the use of wearable sensors for gait monitoring in patients and especially their validation protocols.

However the proliferation of these IoT technologies for remote monitoring implies the production of a huge amount of data, and tends to participate in the increasing quantity of clinical data available electronically.

I.1.1.5 Gait analysis and Machine Learning (ML)

With the generation of sensor networks where each sensor monitors data at different locations, sends data to clouds and generates huge amounts of data sometimes in real-time, this big data configuration needs to use automated process and analysis. ML algorithms appear to be efficient when the size of data produced is extremely large, with a complex interaction between numerous variables. ML can extract valuable information and make useful inferences from data of a multitude of different devices. In Section II.2 we identify over the last decade an increasing number of ML evaluations in research to predict the evolution of chronic diseases. ML aims to build a model that can make repeatable predictions in a high-dimensional space and take into account the non-linearity resulting from the complex relationship between the physical sensors and the classification output. ML analysis can be separated on the following different detection aspects in relation to gait monitoring:

Gait activity detection corresponds to the basic study in gait analysis. It consists of detection from motion sensor data of various gait activities made by the user such as walking, jogging, walking upstairs, walking downstairs and also transition phases such as sit to stand and stand to sit. This study is made both on healthy [20] and pathological groups. For example, Leightley and al. [182] use Support Vector Machine (SVM) and Random Forest (RF) (see Section II.1) to classify ten rehabilitation activities, by extracting the kinematic location, velocity and energy from sensors located on the skeletal joints. Biomechanical studies on gait are useful to identify specific impairments and abnormalities, but it is difficult to analyse complex implicit interactions between many variables in a gait system [258]. Thus, ML models are extremely efficient for the analysis of high-dimensional data with sometimes the number of input variables exceeding the number of samples, and even for healthy and impaired patients. These models are now sufficiently effective to detect different activities in a short amount of time of a few milliseconds so that it can assist clinicians to plan and refine rehabilitation of patient's mobility.

Gait event detection The gait event detection aims at determining the different segments of a step cycle and mainly the heel strike when the foot is in contact with the ground and the toe-off when the foot is not in contact with the ground (see Figure I.2). Other segments of the gait are also analyzed such as stance or swing. Farah and al. [92] evaluate four segments of gait in different walking conditions (*i.e.*, surface levels, walking speeds) using a decision tree (see Section II.1) applied on signals from sensors on the thigh and knee. Twenty specific features were calculated like knee flexion angle or thigh-segment angular velocity. This analysis can provide information on individual variations and dynamic assessment of one's gait. Jung and al. [160] used an exoskeleton for the rehabilitation of stroke patients, which aims at recognizing the intention of the patient before moving, which segment of the gait the patient wants to do. The authors use a gait phase recognizer method based on a neural network trained on hip and knee joint angles and foot status value (whether the foot contacts the ground or not). Yoo and al. [342] study changes in different gait movements to predict the severity and prognosis of knee osteoarthritis in patients using SVM model trained on specific features such as time of stair ascent and kinematic data, including angular features of the pelvis, hip, knee, and ankle. Gait event detection using ML is then beneficial during daily living activities to monitor and detect quantifiable shifts in gait patterns that can lead to clinical interventions specific to each individual.

Gait disorder detection This area aims at detecting gait abnormalities and some deviations from normal walk. One of the interests is the association between neurological conditions and specific gait disorders which is usually the inability to coordinate movements. Before the arrival of ML models, the quantification of the deterioration of a gait was defined with a normalcy index [261, 278], designed to represent and quantify the deviation of a subject's gait from the unimpaired population. Now ML models can determine with high accuracy (>90% depending on the application [179]) abnormalities in gait and the use of multimodal data by integrating with inertial sensors, ground reaction forces or electrophysiological data help models to perform significantly better and seems suitable for the assessment of gait abnormalities [179]. Gait disorder detection has been applied on several neurological diseases such as stroke using RF [68], Hierarchical Weighted Classifier [255] and Neural Network [160], Parkinson's disease using RF, SVM and Neural Network [183], multiple sclerosis or cerebral palsy using k-Nearest Neighbors (k-NN) and Neural Network [13].

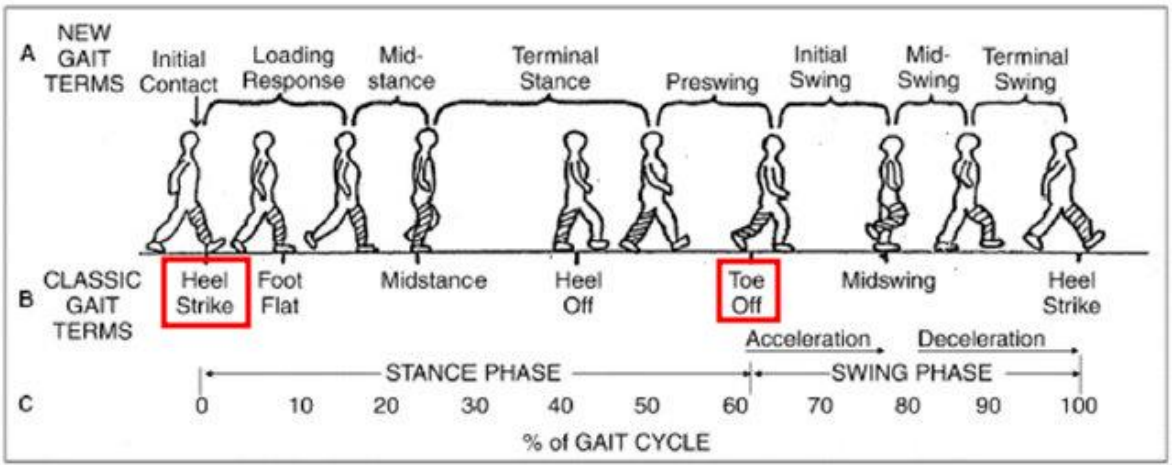


FIGURE I.2 – Illustration of different phases of the gait cycle. The heel strike and the toe off are respectively the starting and ending of the stance phase. *Illustration reproduced from "Towards Effective Non-Invasive Brain-Computer Interfaces Dedicated to Gait Rehabilitation Systems", Castermans and al, 2013, Brain Sciences 4(1):1-48*

Gait asymmetry detection This task concerns the detection of differences in the gait motion of two lower limbs which can result in an asymmetric gait. Symmetry indices are calculated following the difference between left and right sides for a given parameter and dividing the result by the bilateral average. The parameters used include vertical ground reaction forces, plantar pressure distribution, speed and stride frequencies. [267]. This detection is mainly used as an indication of fall-risk for geriatric studies but also for rehabilitation such as abnormal heel strike detection using SVM model [242]. The use of multimodal data with inertial sensor data is also beneficial for this task, Ghasemzadeh and al. [115] used force plates and Electromyography (EMG) signals to improve fall prediction using k-NN and neural network on a set of 28 EMG features and 5 inertial features.

In section II.2, we highlight the contribution of ML for gait assessment and give different recommendations to conduct a study that uses wearable sensors to track patients' gait.

I.1.2 Security and privacy issues

The complex workflow of collected medical data multiplies the security and privacy risks all along the life-cycle of the data including the data collection and transmission [23, 334], as well as the processing and the storage [266]. The concept of privacy means that patients can only reveal information about themselves according to their own choice. In other words, no unwanted information about a patient is revealed to the public. In 2017, the number of attacks on IoT devices has increased by 600% [65] showing a high interest for attackers to compromise IoT systems. With the increased number of IoT devices connected to the Internet (it is projected to increase to 25 billion devices by 2025 [288]).

I.1.2.1 Personal data leakage

The data extraction and analysis from IoT devices may reveal personal information. In 2018, the General Data Protection Regulation (GDPR) was applied and defined a stricter privacy legislation by considering health data as very sensitive personal data. Following GDPR: *"personal data means any information relating to an identified or identifiable natural person ('data subject')* (see Art. 4 (1) of GDPR [1]) and considers data concerning health as : *"personal*

data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status" (see Art. 4 (15) of GDPR [1]).

When such medical data can be accessed by an adversary, multiple risks of privacy threats may occur. For example, leakages of user re-identification are very high (*e.g.*, the re-identification of Governor William Weld's medical information [180]) but also user sensitive information leakage with the exploitation of vulnerabilities in smart home devices [22] and the disclosure of data from wearable fitness tracking devices [353]. Storing and analyzing all raw data on a cloud may also be problematic and raise privacy risks. Communications may be vulnerable to eavesdropping (*e.g.*, Man-in-the-middle attack), attackers only have a unique target to steal data, and more generally moving data to a cloud service provider make user and clinical entity lose control over sensitive data [11]. Moreover the use of ML in healthcare systems has paved the way for numerous attacks on ML models directly (*e.g.*, poisoning attacks [149], membership inference attack [285], attribute inference attack [109]), these aspects are further explored in Section II.3.

However, since the new regulation GDPR, different constraints must be taken into account by the controller who is *"the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data"* (see Art. 4 (7) of GDPR [1]). The controller should be able to demonstrate compliance with a set of six principles including (see Art. 5 of GDPR [1]):

— **Data minimisation** means that personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

— Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject which corresponds to **lawfulness**, **fairness** and **transparency** concepts.

— Data must be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures, which corresponds to **integrity** and **confidentiality** concepts.

— **Purpose limitation** implies that the data are collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.

— **Accuracy** principle involves that personal data shall be accurate and, where necessary, kept up to date.

— **Storage limitation** implies that data must be kept in a form which permits identification of data subjects for no longer than necessary for the purpose for which the personal data are processed.

To apply these principles in practice, national independent agencies in charge of data privacy (such as CNIL in France) give practical tools and appropriate measures to ensure compliance with GDPR [111]. For example, the agency advises to record all the processing activities which allows to make an inventory of the data processing and to have an overview of what is doing with the concerned personal data. The agency also gives some tools to assist in the realisation of Data Protection Impact Assessment (DPIA) which aims at evaluating the characteristics of a data processing, the risks and the measures to adopt.

The GDPR does not include a general obligation of anonymization. It is one solution, among others, to be able to use personal data in compliance with the rights and freedoms of individuals. Anonymization opens up potential for re-use of data that was initially prohibited due to the personal nature of the data used, and thus allows actors to use and share their database without violating the privacy of individuals for example data can be shared

for processing services. It also allows data to be retained beyond their retention period. Therefore, Chapter III mainly focuses on the development of anonymization frameworks.

Even with strict constraints given by GDPR, the risk is never zero. In February 2021, the medical data of around 500,000 people, were stolen from 30 medical laboratories in France [102]. The file contained names of patients, together with their address, telephone number, email and social security number, but also highly confidential information about some of the patients' health, including pregnancies or fertility problems, underlying conditions such as HIV, and medication prescribed. The files were stolen with a fraudulent access to a server associated with the software used by the 30 laboratories to collect their patients' data. In October 2020, the Council of State in France recognized the existence of a risk transfer of personal data from the European Union to the United States in the Health Data Hub [305]. The Health Data Hub is an information system designed to gather all health data of the entire population receiving care in France. As the hosting of the platform has been entrusted to Microsoft, the Court confirmed that on request from the US intelligence services, the risk of health data transmission cannot be excluded.

I.1.2.2 Quantified self democratization

Nevertheless, the production of personal data is not limited to the medical field. Since 2007, the movement of *Quantified Self* [329] has appeared and can be presented as a collaboration between users and tool makers who share an interest in self-knowledge through self-tracking. These practices seem to illustrate a new relationship with the body and data and probably foreshadow new uses linked to the development of connected objects. This movement participates in the development of IoT and aims at producing data at the frontier of well-being and health.

Today there is a wide range of sensors and several thousand applications that mainly concern the following themes: quantifying an activity or a physical parameter (Runkeeper [264], Runtastic [265], Fitbit [98], etc.); monitoring nutrition through calorie estimation (MyFitness Pal [219], etc.); monitor weight (Withings [333], Terraillon [328], etc.); measure sleep quality; monitor a risk factor; assess mood, and so on.

These tools can also differ in the way they record and capture data. In some cases, the data is recorded automatically by a sensor - either external or incorporated into the smartphone - and is then sent back to the editor of the application or sensor [98, 264]. In other cases, the data is entered manually, in a declarative manner, by the user in a dedicated interface such as the application Brightself [47] that assesses and monitors depression and mood during the antenatal period thanks to the completion of momentary questions.

As soon as it is the users themselves who equip themselves to monitor their state of health, outside the medical devices, we leave the traditional framework of medical practice. These measurements carried out outside of a supervision raise several series of concerns especially as wearables are becoming more and more efficient in collecting, with increased accuracy, a number of biomechanical parameters (foot strike pattern, stride length, step rate, etc.) [101] through which to quantify gait and from which applications can derive meaning. These advances in technology make the data produced sometimes as sensitive as medical data. However, the companies associated with these technologies provide less guarantees in terms of security and transparency than a hospital producing medical data using motion sensors for example. In this context, the issue of data sharing becomes more significant, especially when an app or device shares data with any number of third parties.

I.1.2.3 The role of the third parties

Almost all applications collect and send data to third parties companies for different purposes such as marketing analysis. In these exchanges with third parties, and from the perspective of privacy protection, the questions concern the possible reuse of data, their security, and the information and control conferred on users. The study "Mobile health and fitness apps: what are the privacy risks?" conducted in 2013 and revised in 2016 by Privacy Rights Clearinghouse [141], by an American privacy protection association on 43 mobile health and fitness applications, shows that the vast majority of them do not offer sufficient protection to guarantee the privacy of their users' data. In particular, the lack of security of communications (not encrypted), the sharing of personal information with advertisers to generate targeted advertising and the sending of "aggregated data" (data gathered by combining multiple individual-level data) in which users can be re-identified by third parties are highlighted. The same "aggregated data" can also be resold for commercial logic: weight-related measurements may, for example, interest research teams or pharmaceutical laboratories. Commercial companies may also be interested in this data in order to set up points of sale (*i.e.*, a sports shop based on geolocated data from Runkeeper). This data is generally described in privacy policies as "non-identifying" in the sense that it is not directly linked to an individual but remains personal data through the geolocation or socio-demographic information that may be included.

Insurance companies now encourage users to share personal data, because an accurate indication of customer activities and lifestyle choices can help to customise policies. Indeed, the insurance industry is imagining insurance policies where the premium would vary according to the physical activity or lifestyle of individuals because of their impact on the risk covered [41]. In recent years, GAFAM industries also entered the market of e-health. Apple released in 2015 the Apple Watch, a connected watch that takes stock of the state of health and supports the user in his daily life and in his physical activities [306]. In 2014 Google bought the Lift Lab which had marketed the Liftware spoon [190], a high-tech anti-tremor spoon to help patients with Parkinson's disease. These companies collect data on a large scale with sometimes millions of users (the number of Apple Watch users worldwide passed the 100 million mark in 2020) and rely on the personal data that consumers generate with their services or products to create personalised services or offer targeted advertising space to other companies. That big companies (GAFAM) are collecting medical data on a large scale raises a number of questions.

I.1.2.4 Data transparency

Once the data has been collected, the provider can only be trusted to process the data according to the stated purpose. However, beyond the obvious lack of transparency for individuals on how the data are used, the study "Mobile health and fitness apps what are the privacy risks?" [141] denounces the discrepancy between the privacy policies presented to the user and the reality of the practices concerning personal data. For those applications that do publish their privacy policy, the authors of the study found that the majority of potentially privacy-risky practices were not described in a clear and understandable way. The authors even explain that they have identified *"a correlation between the level of detail of a privacy policy and the risk in the use of the application regarding privacy"*. They add: *"the more detailed an application's privacy policy was, the more privacy-invasive the practices found were"*. This makes it difficult for users to have real control over their personal data. To strengthen ethics, ensure the protection of the rights, and regulate the practices in companies some ethics committees have been created, but they are often influenced by conflicts of interest [307].

GDPR has now defined different types of information that needs to be provided to the data subjects such as identity of the controller, purpose of the processing, etc. The recital 58 gives the following requirements : *"The principle of transparency requires that any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used"*.

However, Morel [216] points out four characteristics of IoT devices that makes information and consent difficult in practice. Their *ubiquity* which means they have invaded our daily-life, their invisibility combined with their proliferation reinforces the difficulty of being informed. Their *variety* means they do not collect the same type of data, do not communicate with the same protocols and the variety of usage contexts makes it difficult to retrieve information. Their *low computational power* means they are fit for a limited number of tasks which generally consists of collecting data and sending it to a central node. Lastly, they have *inappropriate interface*. Due to their small size and the fact that they are battery-powered, some devices literally do not have the necessary means of communication to interact with the user.

I.1.2.5 Security and privacy threats

According to Nokia Threat Intelligence Report 2020 [230] IoT infections has increased by 100%, IoT devices are now responsible for 32.72% of all infections observed in mobile networks, up from 16.17% in 2019. In practice, the production of data using self-quantified applications is exposed to numerous attacks at different times of the processing data framework. Starting from the data acquisition, IoT themselves are prone to many attacks, we introduce here only some of them. The most obvious threat is the use of weak or guessable passwords. Lots of IoT devices, specifically those that use web interfaces, are not always reconfigured to let the user change the default password which leaves the device vulnerable to brute-force attack for example [295]. The code-injection attack aims at introducing malicious code into the system thanks to errors in the program [93, 213]. This attack can be used to steal data, get the control of the system or even propagate worms [337, 351]. Spoofing attacks can be easy to implement in an IoT access network by using a faked identity such as the Media Access Control (MAC) or Internet Protocol (IP) address of the real user. The attacker then can claim to be a legitimate device and can have access to the IoT network [324]. Indeed, at the network level which concerns connectivity of an IoT system, the transmission medium is often wireless with Bluetooth, WiFi, 3G, etc. One of the most common attacks is Denial-of-Service (DoS) attack which consists of flooding the targeted machine with requests. As a result of sudden incoming traffic, the online service can be unavailable to users due to resources exhaustion [166]. This attack can be used as a smokescreen to achieve other attacks to violate the defensive system and therefore the privacy of the user's data [93]. Man-in-the-middle attack is also a common attack that occurs when an attacker intercepts the communication between two systems in order to obtain information of the two entities [226]. At the cloud level of IoT systems, among the most common attacks the malware injection attack aims at introducing a malicious service or a virtual machine into the cloud in order to redirect user's requests to the malicious module and execute the malicious code. Then the attacker can manipulate or steal data with an eavesdropping attack [127]. With billions of records stolen worldwide every year [2], stolen data often ends up being sold online on blacks markets. For example, in 2018 hackers offered for sale more than 200 million records containing the personal information of Chinese individuals [128].

Beyond security issues, with the high amount of data produced with IoT, the data controller is not always able to ensure that user personal information is hidden and then vulnerable to improper sharing and misuse, which makes privacy also a major concern. It is

known that data from GPS, cameras or microphones are highly sensitive for user's privacy, but motion sensors are also concerned. Data extracted from motion sensors can be used for re-identification [178], location tracking [130], behavior tracking [201], keystroke inference [51], demographics information inference [211], psychological traits inference [352], etc. Furthermore the processing framework is also concerned by privacy issues especially when ML models are used. Indeed, these models are able information in their parameters not related to the task targeted and then may intentionally contain sensitive information about the user's data [347]. These aspects are discussed in more detail in chapter II.3.

I.1.3 Research problematics

More reliable and secure alternatives regarding the collection and processing of healthcare data and specifically motion sensor data needs to be explored. In the context of activity recognition via wearable sensors, the challenge is then to identify ML methods that can preserve the privacy of individuals while maintaining sufficient relevant data for ML tasks [290]. This challenge raises two important questions:

- Is the collected data sufficiently protected so that no one can misuse it to infer sensitive information or re-identify the owner?
- How do we determine if the protected data is still accurate enough for healthcare applications such as rehabilitation?

Achieving this balance between data utility and privacy is an important goal for sending secure and trusted data via wearable sensors and building end-user trust and adoption.

I.2 Fields of investigation and contributions

To answer these two questions, my thesis aims at proposing in the different contributions, a paradigm shift from a centralized strategy where the preprocessing, the storage and the analysis of data are done on a single location, to a strategy that move a part of the processing locally (*i.e.*, on the user's smartphone). These local processes aims at identifying the relevant information for the gait monitoring (sent to the server) from information leading to a leak of sensitive information (not disclosed). To assess the feasibility of this approach, we also evaluate the cost of operating these minimization schemes on user's devices.

We detail in this section my five main contributions. The first one is a systematic review of the validation of commercial wearable sensors for gait monitoring in healthcare that highlights the contribution of ML over the last decade (see I.2.1). The second contribution explores different aspects of privacy preserving ML, firstly on the anonymization of data through minimization and thus limiting the collection and storage of data that allow the re-identification of patient (see I.2.2 and I.2.3). Then the protection of other sensitive attributes using Generative Adversarial Networks (GAN) (see I.2.4). Finally we explore the protection of sensitive attributes on data stored in a decentralized manner (see I.2.5).

I.2.1 Contribution of ML in validation of wearable sensors

Previous research has shown significant differences in spatiotemporal gait parameters between similar in-lab and in-field studies, illustrating the importance of establishing commercial sensor validity for long-term patient monitoring. Indeed, clinicians are still cautious to use due to their doubts about the quality of the data collected as well as the reliability of the sensors. In this context commercial sensors must be validated with rigorous validation methods.

In chapter II.2, the scoping review summarizes the state of the art between 2010 and 2020 in terms of the use of commercial wearable devices for gait monitoring in patients. For this specific period, ten databases were searched and 564 records were retrieved from the associated search. This scoping review included 70 studies investigating one or more wearable sensors used to automatically track patient gait in the field.

This contribution explains why studies using ML are tending to become more numerous. ML brings benefits compared to statistical validation methods, as the large amount of data production makes ML methods robust, efficient and fast to analyse a complex and high-dimensional data space. ML is particularly adapted to predict the evolution of a disease and the corresponding rehabilitation.

I.2.2 Data minimization through local pre-processing

Motion sensor data are usually transmitted to analytic applications hosted in the cloud that use ML models to perform activity recognition but this data also contains private information about users without their awareness and may even cause their re-identification.

In chapter III.1, I propose a framework to efficiently recognise the user activity, useful for personal healthcare monitoring, while limiting the risk of users re-identification from a set of features calculated based on accelerometer and gyroscope signals. This framework relies on local processing to minimize the data transmitted to the server. To achieve that, we show that features in the temporal domain are useful to discriminate user activity while features in the frequency domain lead to distinguishing the user identity. This protection mechanism

extracts the most important features for activity detection and limits the ability to re-identify. These unlinkable features are then transferred to the cloud. We extensively evaluate our framework with reference datasets. Results show an accurate activity recognition while limiting the re-identification rate.

I.2.3 Data anonymization based on time-frequency representation

We also explore a novel representation of motion sensor data in two dimensions. Time-frequency representation gives the frequency evolution of the signal components as a function of time. In chapter III.2, I propose a privacy-preserving framework for activity recognition based on this representation in order to deal with the non-stationarity of the signals and therefore allow a better trade-off between activity recognition and user identification.

Acceleration signals (x, y and z axis) are encoded in the time-frequency domain by three different linear transforms: Short Time Fourier transform, Stockwell transform and Optimized Stockwell transform. Second, we propose a method to anonymize the acceleration signals by filtering in the time-frequency domain. Finally, we evaluate our approach for the three different linear transforms with a convolutional neural network classifier adapted to image processing, with a late fusion strategy of x, y and z axis images.

I.2.4 Data sanitizing to prevent inference of sensitive attributes

Anonymization is not always sufficient, the disclosure of data may also lead to sensitive attribute processing not consented by the data subject and sometimes discriminating treatment based on social or physiological information between the users. To address this issue, we propose in chapter IV.1 DySan, a privacy-preserving framework to sanitize motion sensor data against unwanted sensitive inferences while limiting the loss of accuracy on the physical activity monitoring. Our approach is inspired from the framework of GAN to sanitize the sensor data.

DySan builds various sanitizing models, characterized by different sets of hyperparameters and dynamically selects on the smartphone the model which provides the best utility and privacy trade-off according to the incoming data.

I.2.5 Federated Learning with personalized layers

Lastly, we briefly explore a distributed learning scheme in chapter IV.2. Contrary to the traditional centralized approach, we specifically use Federated Learning (FL) to train a learning model across multiple participants without explicitly sharing data samples.

While FL is a clear step forward enforcing users' privacy, different inference attacks are still possible. By only sharing learning models and not data, the framework is exposed to different attacks such as inferring the user's sensitive attribute based on the model's parameters.

In this chapter IV.2, we quantify the utility and privacy trade-off of a specific FL scheme. Instead of each user sharing their entire model, only a part of the model is shared (called *upper layers*) for the FL and the rest of the model is kept (called *personalized layers*) for the local training. While this scheme has been proposed as local adaptation to improve the accuracy of the model through local personalization, it has also the advantage of minimizing the information about the model exchanged with the server. However, the privacy of such a scheme has never been quantified.

Our evaluations using motion sensor dataset show that personalized layers speedup the convergence of the model and slightly improve the accuracy for all users compared to

a standard FL scheme while better preventing both attribute and membership inferences compared to a FL scheme using local differential privacy.

Chapter II

Background and Related Work

This chapter aims at exploring in the literature the two main aspects of this thesis which firstly consists of using ML on complex and large-scale motion data to assess rehabilitation and secondly also using privacy preserving ML scheme to overcome privacy issues raised by the production of motion data.

The first section gives a short overview without giving mathematical details of what ML is and presents several methods used in this thesis. The second section further explores the benefits of ML to validate the use of motion sensors for gait monitoring. We conducted a review of the validation methods used by studies in healthcare rehabilitation over the last decade. The third section focuses on the privacy issues that the large production of motion sensor data implies and specifically the attacks that aim at inferring sensitive information on data acquired then, we explore the main defences designed to overcome these attacks.

II.1 Short overview of ML

ML is an automatic method for data analysis that is a revolution in many science aspects and will significantly influence the research in healthcare in the near future. The hype for methods around ML is already happening with more and more publications on the subject. This tendency can be observed by looking at the number of papers published over the years and especially in Pubmed, a database for biomedical field (Figure II.1)

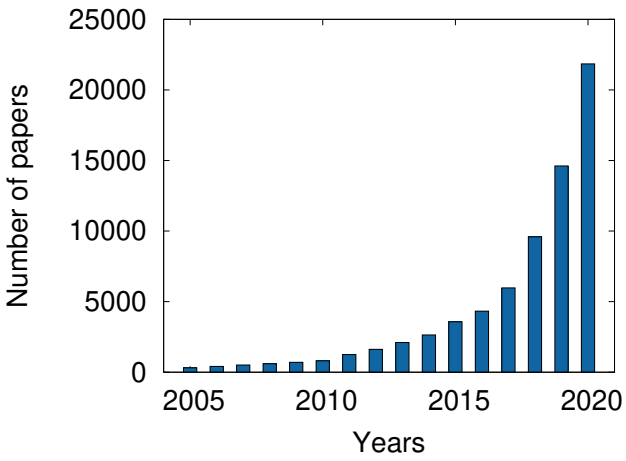


FIGURE II.1 – Number of papers published in Pubmed.com using the search term (ML) OR (deep learning) and choosing a specific year in advanced search. Pubmed is a database for biomedical field.

II.1.1 A two step process

In general, ML models aim at automatically inferring general relationships on large datasets for complex tasks that are usually not possible to make by hand. The main way

to build a ML model relies on 2 steps:

Training phase. By considering a training dataset as a combination between the input data $X = x_1; x_n$ (with n the number of training examples and x_i the feature vector) and $Y = y_1; y_n$ the output space that determines the task that the model wants to infer. Then, the model is included in a space of functions $\theta, X : f_\theta(X)$ which corresponds to a set of functions that models the distribution that represents the dataset. The objective of the training phase is to find the set of parameters θ that best fits the model prediction with the output data Y. This is done through a loss function $L(f, x, y)$ that quantifies the differences between the model's prediction and the true values. Concerning neural networks, the training often consists of minimizing the loss function thanks to the Stochastic Gradient Descent (SGD) method which consists of iteratively calculating the function f and the loss l for each $x \in X$. The model is evaluated on a validation dataset disjoint from the training dataset to optimize the hyperparameters of the model.

Test phase. After the training phase, the model can be used on a test dataset disjoint from the training and validation dataset in order to make predictions and measure how well the model generalizes. These predictions are generally represented by a vector of size the number of classes in the output space. Each element of the vector corresponds to the probability that an input data belongs to the class.

II.1.2 Three classes of tasks

The ML tasks are generally divided into the three following main classes:

Supervised learning. In this configuration, the input data are labelled with a corresponding output data. The goal of the model is to map the input data to the output labels even for new input data unseen by the model during the training. If the output data is categorical, we call the task made by the model classification. If the output data is continuous, we call the task made by the model regression. Our contributions are in majority based on the classification configuration.

Unsupervised learning. This configuration happens when the input data is not labelled with a corresponding output data. It generally concerns clustering methods such as k-means [221] or dimensionality reduction methods such as Principal Component Analysis (PCA) [156]. A clusturing task consists of grouping a set a data in such a way that the data in a group are more similar to each other than the data in the other groups. This similarity is generally measured with a distance function.

Reinforcement learning. A model built in this configuration does not need input/output data either but the model learns how an agent can take decisions in an environment with rewards. The agent learns by exploring the environment and experimenting with decisions. Generally, the learning process is based on an objective which consists of maximising the cumulative rewards during the repeated trial-and-error interactions with the environment [181].

II.1.3 From shallow ML to deep learning

ML domain is composed of numerous methods. In this section, we present different common methods used along the thesis. We differentiate *shallow models* and *deep learning*

models based on neural networks. The first category is generally based on preprocessed data to generate new handcrafted features based on the raw data. The identification of effective features depends on the application requirements, experience in the research field, or the prior domain knowledge provided by experts. In gait analysis from motion sensor data, gait features are categorized according to 3 levels: (i) *low* where the analysis is done on raw signals without post-processing, (ii) *medium* where the analysis is based on statistical descriptors extracted from the signals (mainly statistical moments or common signal processing features) and (iii) *high* where the analysis is based on descriptors at a high level of representation which disregard the technical characteristics of the equipment or methods used (e.g. step length, cadence and number of steps). Concerning shallow ML, medium features are generally calculated. Unlike deep learning methods which automatically learns effective features from the raw data thanks to the different layers in the neural network. Several shallow models are used in the different contributions:

Decision tree is a non-parametric (*i.e.*, no distributional assumptions is made on the data) supervised learning method that can be used for both classification and regression tasks [225]. The building of decision tree involves several steps. The splitting of the input data into subsets from the root to the leaf nodes. Each node is a decision based on a feature that split the dataset into two subsets. The process of splitting is called Attribute Selective Measure (ASM) based on the calculation of entropy which measure the homogeneity of the subsets. We calculate the information gain based on the decrease of entropy after the dataset split, and choose the feature with the largest information gain as the decision node. The entropy is calculated with the following formula:

$$E(S) = 1 - \sum_{i=1}^{c} -p_i log_2 p_i$$

with S the subset, i the class of the data and p_i the probability of a data point to belong to the class i. A last step called pruning aims at removing parts of the tree to avoid overfitting which means that the model learns too much details and even noise in the training data so that the performance on new data is negatively impacted.

Random Forest (RF) is an ensemble learning model, an approach that aims to combine the predictions from multiple models to increase predictive performance. RF is built with a multitude of decision trees. The model is built with a Bootstrap Aggregating (bagging) method where data is sampled and distributed to the different decision trees, the different predictions are then averaged resulting in better performance than any single tree in the model. When the amount of data is not significant, RF is one of the most efficient choices for human activity recognition [77], for this reason among others RF is applied in Section III.1.

Support Vector Machine (SVM) is an algorithm that define a space in which the different classes are maximally separable. SVM finds a hyperplane characterized by support vectors (*i.e*, data observations that determine the decision boundary) that maximizes the margin separating the classes. linear-SVM finds the hyperplane in the feature space of the input data. When the data are non linearly separable in the input space, nonlinear-SVM transforms the input feature space into a higher-dimensional space, based on a kernel function defined by the user, where the data are now linearly separable. The decision boundary will be then a hyperplane in this higher dimensional space.

K-Nearest Neighbors (k-NN) is also based on distance metric to classify data samples into different classes. We consider a set of data in a multidimensional space and k a constant defined by the user. To classify a new unlabeled data, the k-NN method assign the label which is most frequent among the k data samples nearest to that unlabeled data.

Naive Bayes classifier is a collection of probabilistic classifiers based on the Bayes theorem. Those classifiers consider the features as independent and calculate the probability of a class output Y given the features X. The input data is classified following the class that gives the highest probability.

Deep learning This subset of ML is composed by multi-layers neural networks that gradually become the most used approach in ML and outperform previous ML algorithms in many domains and is now commonly used for gait monitoring [14]. Deep learning approaches are particularly efficient when the size of the data produced is extremely large, which requires discovering hidden patterns in the data, a deep understanding of relationships between a large number of interdependent variables and a solution adapted to specific cases which may update over time. Unlike the methods presented previously, specific features are not necessary to be designed, the features are calculated automatically by the model and optimized for the task considered.

We will further explain different deep learning methods used in this thesis, the overview is not exhaustive and algorithms such as Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) are not presented.

Fully connected neural networks is one of the most common neural networks composed of multiple layers. The input layer is composed with the input data, the hidden layers are composed with artificial neurons defined by their weights w, bias b and an activation function f that aims at transforming the data received by the previous layers in order to send it to the next layer. There are numerous activation functions (sigmoid, hyperbolic tangent, Rectified Linear Unit, etc.) that perform a non-linear transformation essential to produce decision boundaries for nonlinear datasets. The activation function may differ for each problem statement.

The output of a layer j is defined by the following equation:

$$o_j = f(\sum_{i=1}^{n} w_i x_i + b)$$

with n the number of neurons in the layer. The architecture of a fully connected neural network is such that all the neurons of a layer are connected to the neurons in the next layer.

Convolutional neural networks (CNN) has the possibility to exploit local correlation in the data by constraining a local pattern between neurons in adjacent layers. It is particularly used for 2D representation data such as images and even 1D representation data such as time series (*i.e.*, motion sensor data). The model aims at representing in feature maps the data with low level features such as peaks in 1D, contours and curves in 2D. As you go through the layers, the representations become more abstract. Figure II.2 represents a simple CNN model for image classification.

The convolutional layer is composed of filters that independently perform a convolution across the width and height of the image. The convolution simply computes a dot product between the filter and the input data. The learning process consists of having filters that activate when it detects specific patterns in the features at some position in the input data.

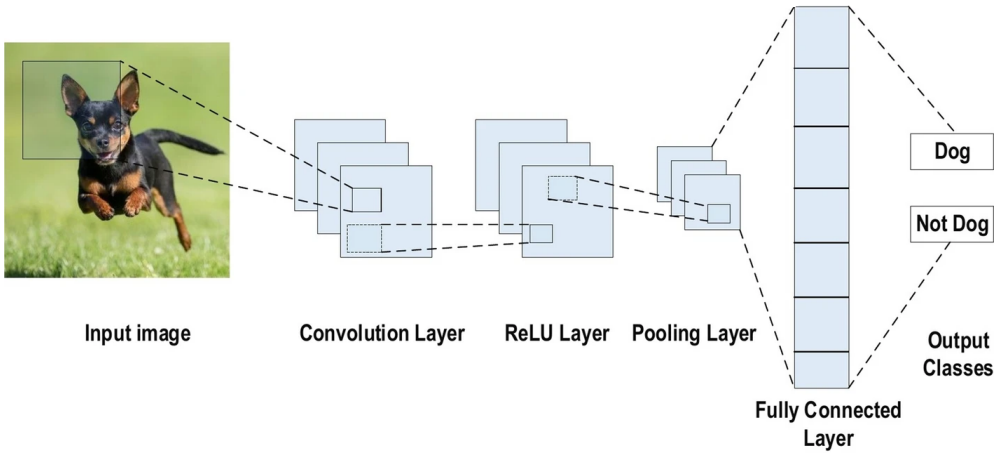


FIGURE II.2 – An example of CNN architecture for image classification. *Illustration reproduced from "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions" Alzubaidi and al. 2021.*

The process is defined by 3 parameters: the size of the filter, the stride corresponding to the number of data points the filter moves for the next position, with a value larger than 1 the feature map dimension will be downsized compared to the input data. The zero-padding parameter aims at specifying the number of zeros to pad around the border of the input in order to preserve the dimension after the process (see Figure II.3).

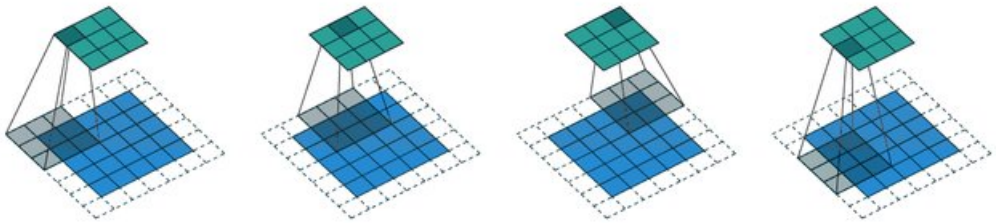


FIGURE II.3 – An example of convolution process with a stride of 2, a (3x3) filter size and padding.*Illustration reproduced from "Deep Learning Operators Optimization in Tiramisu (Sparse Neural Networks and Recurrent Neural Networks)" Debbagh and al. 2020.*

After a convolutional layer we usually apply a Rectified Linear Unit (ReLU) activation function ($f = max(x, O)$) to increase the non-linearity in the data, for an image it extracts shapes and contours in the feature map. Finally the pooling layer downsample the feature map by summarizing the presence of features in patches of the feature map. As the convolutional layer, it consists of a window that scans the input data with a window size, stride and zero-padding parameters. At each time the window is moved, a pooling function is applied. The most common is max-pooling which extracts the maximum value of the window. To further model non-linear relationships in the feature map, a fully connected layer is sometimes added at the end of the network before the output layer.

Autoencoder is a typical unsupervised ML algorithm based on two distinct parts represented in Figure II.4. The Encoder is a set of layers parameterized by weights and bias which can be either fully connected or convolutional layers. This part compresses the input data into a lower dimensional *code* also called the *latent-space representation*. The Activation

is a nonlinear function that transforms the encoded data. The Decoder is a set of reverse layers that produces the reconstruction of the data into the same space representation as input data. This algorithm has numerous applications as data denoising, data synthesizing and so on [175]

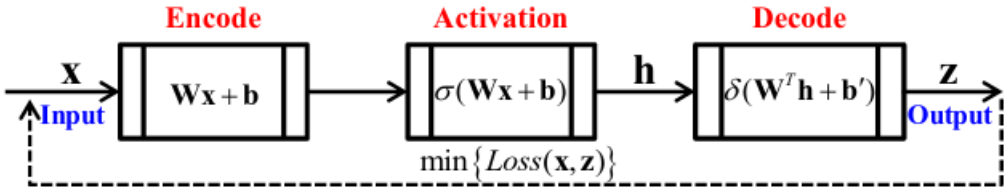


FIGURE II.4 – The generic flowchart of autoencoder. *Illustration reproduced from "A Review of the Autoencoder and Its Variants: A Comparative Perspective from Target Recognition in Synthetic-Aperture Radar Images" Dong and al. 2018.*

Generative Adversarial Networks (GAN) is a ML framework used in Section IV.1. It was designed by Goodfellow and al. [120] and consists of a competition between two models in zero-sum game where the two models are simultaneously trained and the gain of a model corresponds to the loss of the other. Figure II.5 represents an architecture of a GAN. The first model called generative model G aims at generating data $G(z)$ from input random noise z (usually from an uniform or normal distribution) by capturing a specific data distribution X. The discriminative model D evaluates the output data given by G by estimating the probability that a sample of G(z) comes from the training data X rather than G. The objective function of G aims at maximising the probability of D making a mistake. The idea is that G is trained to fool D by producing data that D perceive is not synthesized.

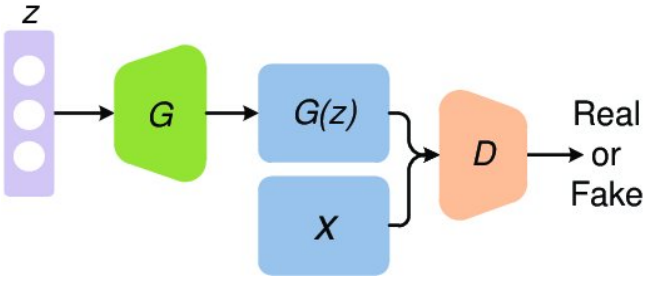


FIGURE II.5 – An example of GAN architecture. *Illustration reproduced from "Recent Progress on Generative Adversarial Networks (GANs): A Survey" Zhaoqing and al. 2019.*

II.1.4 Centralized versus distributed learning

Usual learning process uses a centralized setting where the data and the model are located in the same place. Even if there are multiple data owners (multiple patients that produce data), their data are collected in one central server where the model can be trained using the full dataset. In a distributed setting, the data of the users stays locally on its device and only a learning model is exchanged with the server which corresponds to a collaborative or FL: iteratively, the server sends a model to devices, this model is trained and refined with the local data on each device and all local models are sent back to the server which

aggregates them into a single model (such as the average of the weights of the local models following McMahan and al. [206]).

At each learning round i, each client k trains its local model m_k with its own data using SGD during several iterations j. In its synchronous version, once all the participants send their model update m to the server, the server then aggregates all these model updates using the following equations before to disseminate back this aggregated model to all devices:

$$M_{i+1} = \sum_{c=1}^{C} \frac{n_c}{n} m_c^{i+1},$$

with n_c the set of indexes of all the data points n on client c, m_c^{i+1} the local update of a client c, computed as follow:

$$m_c^{i+1} = m_c^i - \eta g_c^i,$$

with η a fixed learning rate (*i.e.*, hyperparameter which controls the step size of the SGD optimization) for each client and g_c^i the average gradient on the local data of the client c at the epoch i. Those learning rounds continue until the convergence of the central model.

FL architecture is further explored in Section IV.2. The process can also be fully decentralized or peer-to-peer (P2P), in these settings there is no central server that manages and aggregates the local models, the users communicate and exchange their models directly with other users, which can be interesting for privacy perspectives because the need to trust a central server is not required [36].

The contributions presented along the chapters of this thesis can be considered as falling between centralized and distributed architectures, because even if in each framework part of the data is sent to a cloud, an important part of the processing and the different privacy-preserving framework are applied locally.

II.2 ML for gait monitoring in healthcare

After giving an overview of the different ML methods, we can review the benefits of ML in validation methods of commercial wearable sensors in gait monitoring for healthcare. There are already many reviews on validation of commercial wearable sensors available in the literature, most were interested in monitoring activity on healthy subjects [80, 89, 94, 172], while others have taken a descriptive approach centered on a very specific medical application [247, 298, 319]. However, few studies focus on the validation methods, the ground truth used and how the reference data are annotated. A common validation method is to use inferential statistics, such as a regression analysis to explore and model the relationship between sensor and ground truth data. These approaches typically assume that the relationship between sensor and ground truth data follows a linear pattern. Linear regression has the advantage of being simple to use and to interpret. In comparison with these linear methods, the nonlinear methods can fit more types of data in terms of shape and are hence recognized to be more general. Some non-linear approaches such as Deep Learning methods have the advantage to be less dependent on the assumption of the model and very recently produced promising results in sensor validation [193, 281]. Non-linearity seems particularly interesting in terms of patient monitoring in order to integrate networks of several sensors placed at different places on the patient [38, 250] and for high-level tasks (such as the classification into groups of patients according to the evolution of a disease) [15, 231] which requires the integration of various information on locomotion and control systems involved in the complex gait regulation [121, 245].

In this section, our aim was to conduct a systematic review to i) determine the statistical methods currently used for the validation of sensors and, ii) determine to what extent ML is used as a statistical method for this validation step. Our expectation is that ML-based methods may provide better validation results since they are able to model more complex boundaries.

II.2.1 Methodology

II.2.1.1 Databases

We conducted a literature search of the PubMed, SCOPUS, ScienceDirect, Web of Science, IEEE Xplore, ACM Digital Library, Collection of Computer Science Bibliographies, Cochrane Library, DBLB, and Google Scholar (first 50 results) databases for all literature published between 2010 and 2020.

II.2.1.2 Literature search

The literature search strategy included a combination of keywords to identify articles that addressed (i) gait assessment/detection, (ii) wearable and connected technology, (iii) chronic pathology monitoring and (iv) validation. Keywords included "gait", "walk", "actigraphy", "actimetry"; "smartphone", "wearable", "mobile device", "IoT"; "chronic disease", "rehabilitation", "medicine"; "validity", "validation", "reliability", "reproductibility". The full search term strategy that was used for each database is given in Table 1 of Appendix.

After an initial screening, which consisted of reviewing all article titles and abstracts, the full content of 102 of these articles was screened in more detail for eligibility. After removing the articles that did not meet the inclusion criteria, 70 articles were deemed eligible for the review.

We now analyze the selected papers by categorising them following different criteria in order to extract common patterns and trends.

II.2.2 Clinical context

The sample size of the studies ranged from 1 to 130 participants, with a mean of 37.89 participants (SD = 30.68) per study. The length of time for data collection in 2 different conditions (laboratory or free living) varied and was not always reported with an exact numerical value or unit. Therefore, we only report in Table II.1 ranges of acquisition times which go from hour to year. Among the selected studies, 33% (N = 25) focused on neurodegenerative diseases [8, 9, 18, 24, 54, 71, 85, 87, 96, 99, 108, 125, 145, 168, 171, 191, 205, 208, 212, 252, 269, 280, 283, 313, 346], 24% (N = 18) on orthopedic disorders [5, 62, 74, 106, 148, 169, 173, 184, 185, 202, 220, 243, 263, 277, 302, 303, 314, 322], 24% (N = 18) on diseases of vascular origin [29, 52, 59, 63, 64, 73, 74, 76, 133, 146, 154, 158, 159, 256, 277, 291, 314, 335], 8% (N = 6) on aging and associated pathologies [21, 91, 153, 259, 277, 315], and 4% (N = 3) on diseases associated with poor lifestyle [48, 133, 176]. Finally, 5 studies were classified as "others" [33, 60, 218, 227, 279] because they could not be grouped together in an existing group.

Acquisition Time	t<1 h	1≤t<24 h	1≤t<7 d	1≤t<4 w	1≤t<12 m	t≥1 y
Laboratory (N = 53)	**46**	3	0	1	2	1
Free Living (N =17)	1	1	1	**8**	3	3

TABLE II.1 – Frequency of studies according to conditions of data collection (laboratory or free living) and acquisition time *t* (from a few minutes to more than a year). In bold is shown the most common acquisition time for each data collection condition.

II.2.3 Wearable sensor types

The most frequently used type of wearable device is the Inertial Measurement Unit (IMU; N = 39) then almost equally the smartphone (N = 18) and a single sensor (N = 17) . The majority of studies (N = 56) used multi-sensor systems (incorporating more than 1 sensor) to automatically assess gait in chronic pathologies. On average, 5.78 wearable sensors (SD = 8.43) were used in the studies, with a range of 1 to 64 sensors (Table 2). As depicted in Table II.2, the most utilized sensor was accelerometer (95 %) either by itself (N = 17) or embedded into a device (N = 57). The second most frequently used sensor was gyroscope (51 %) followed by magnetometer (14 %) and others (16%).

II.2.4 Data acquisition conditions

Most of the papers collected their data in laboratory conditions (N = 53) while a smaller part did in free living conditions (N = 17) (see Table II.1).

Regarding the positioning of sensors and/or devices (Table II.3), 60% of the studies place them on the inferior part of the body , generally on the feet (N = 14) or on the hips (N = 6). Chest location is also widely used (49%) ; 17% of the studies have carried out sensors positioning on hands and arms (superior) while the other 17% used a trouser or jacket pocket

Device type		Sensor type	
IMU	**39**	Accelerometer	**39** (100%)
		Gyroscope	30 (77%)
		Magnetometer	8 (20%)
		Others	7 (18%)
Sensors	17	Accelerometer	**14** (82%)
		Gyroscope	1 (0.7%)
		Magnetometer	1 (0.7%)
		Others	4 (3%)
Smartphones	18	Accelerometer	**17** (94%)
		Gyroscope	7 (38%)
		Magnetometer	2 (11%)
		Others	1 (5%)

TABLE II.2 – Frequency of devices and sensor types in included studies. The device is the tracker used by the patient (first column), which may include different sensors which are detailed in the second column. Note that since a device can use several sensors, the total number of occurrences in the second column is much greater than that of the first column.

Superior	Inferior	Chest	Free
12	42	34	12

TABLE II.3 – Frequency of sensor locations reported on the patient from included studies.

II.2.5 Gait indicators

A wide majority (70 %) of studies (see Table II.4) use high level features for gait analysis which can be correlated to the high use of smartphones (in the studies reviewed, see Table II.2) that already compute this type of features on the device.

A significative part of studies (28 %) use medium level features while low level features (raw data) are much less exploited (8 %) .

Low level		Medium level		High level	
Total	**6**	Total	**20**	Total	**49**
		Magnitude mean	11	Step length	20
		Magnitude standard deviation	10	Number of steps	18
		Peak frequency	9	Cadence	15
		Mean crossing rate	5	Speed	11

TABLE II.4 – Frequency of features extracted from sensor signal reported from included studies. These different features were classified into three categories described in section II.1.3.

II.2.6 Ground truth

Ground-truth methods are categorized according to 6 levels: (i) *controls* where a group of subjects serves as a reference, (ii) *expert* where the data is analyzed with regard to annotations made by experts, (iii) *med device* where the data is analyzed with regard to a portable device already used in clinical routine, (iv) *medical* where the data is analyzed with regard

to a medical examination/test or clinical score, (v) *metrologic* where other high resolution equipment is used as reference and (vi) *user annotations* where the data is analyzed with regard to annotations made by patients during use of the device.

To evaluate the validity of the commercial wearable sensors for gait monitoring in patients, all the studies (N = 70) used one or more validation methods in which "ground truth" data was reported. About half of the studies (53.3%) use annotations and the other half (46,7%) a reference to validate the results from the sensors. Regarding annotations, most studies use labeling according to 2 or more groups of subjects (the vast majority of the time a group of patients and healthy controls) , others use annotations made by experts on data from videos or measurements during the experiment and finally four studies had participants self-report via log or diary. As regards the reference to which the studies compare the data from the sensors, it concerns in equal parts a metrological device (18.3%) or a medical examination (20.2%) and to a lesser extent (8.3%) a third-party portable medical device

II.2.7 Evaluation methods and metrics

Evaluations methods are categorized according to 5 levels: (i) *descriptive stat* where evaluation is carried out through descriptive statistics only, (ii) *descriptive stat + test* where evaluation is carried out through descriptive statistics with statistical tests, (iii) *linear models + stat test* where evaluation is carried out through linear models with statistical tests, (iv) *ML* where evaluation is carried out through ML only and (v) *ML + stat test* where evaluation is carried out through ML with statistical tests.

Studies often reported multiple and varied evaluation metrics. All reported evaluation outcomes and their corresponding evaluation method are included in Table 3 and depicted in Figure II.6. The most common evaluation method was descriptive statistics (61.4%) including or not statistical tests where correlations, mean errors or p-values are most commonly reported. The other evaluation methods go through modeling either by a linear model (11.4%) or a ML model (17.2%) . Due to the lack of a standardized evaluation metric across studies, we do not summarize (calculate mean, standard deviation, etc.) the reported metrics. However, evaluation metric values – as given in the abstract or the conclusion of the associated studies – are available in Table 3.

A closer look to studies using ML highlights that ML-based approaches are often used for high-level validation tasks (see Table II.5), such as distinguishing between different groups of patients or stages of disease progression [5, 8, 9, 59, 158, 168, 208, 252, 303]. This is an important point, because ML aims at generalizing a model to patients not included in the initial dataset. Another point to emphasize, as illustrated in Table 5, is that studies using ML as a validation method incorporate a large number of variables (the complete raw signal or a collection of different sensors) [5, 108, 145, 168, 191, 208, 212]. This is not the case in studies using statistical methods which work with a few dozen of variables at the maximum and often in a univariate way two by two [18, 91, 106, 269, 335, 346].

Number of studies	<10	10-100	>100
Statistical	43	8	0
ML	3	9	7

TABLE II.5 – Frequency of studies using respectively less than 10 descriptors, between 10 and 100 descriptors and more than 100 descriptors for the validation on both statistical and ML methods

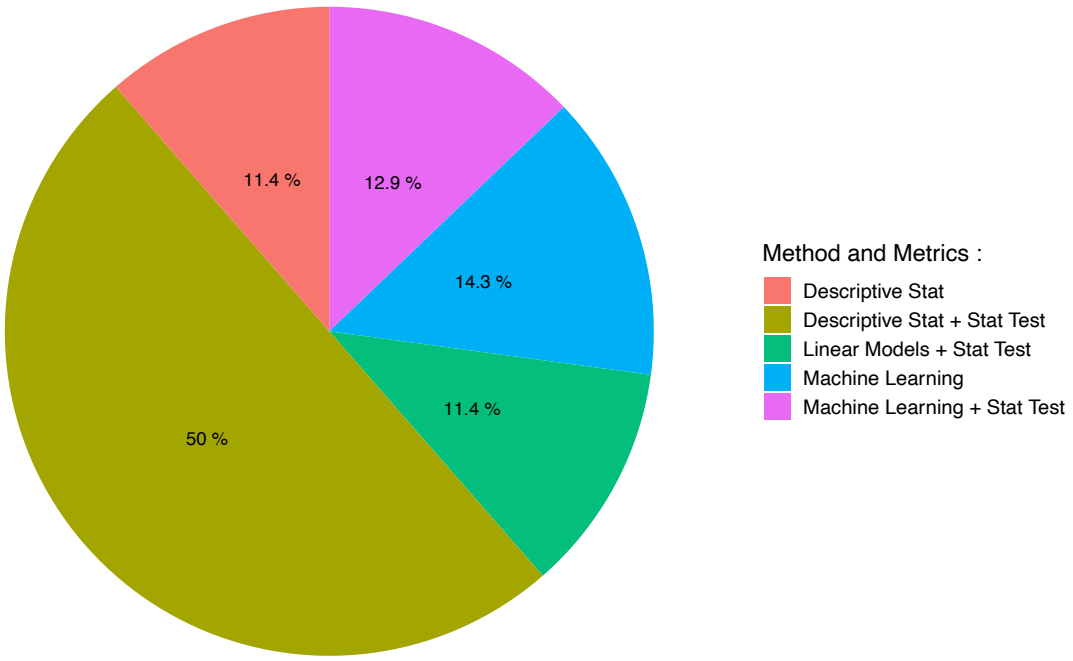


FIGURE II.6 – Pie chart representing the percentage of papers using the different levels of evaluation identified among the 70 selected papers.

II.2.7.1 Trends and challenges

Acquisition context. Most of the first studies were restricted to the laboratory environment and over short acquisition times (of the order of a few minutes). The first papers to report sensor validation in a free living environment were in 2011 [76, 176]. As seen in Table 2, from 2017, studies of this type become more frequent [60, 71, 73, 74, 87, 106, 133, 153, 191, 252, 283, 302, 313, 346] due to changes in the sensors which are detailed in the following section.

In this review, we also observe that early research efforts attempted to find improvement for gait monitoring in patients by experimenting with new sensor types and/or sensor locations. Over time, research efforts have focused on refining validation protocols, whether in terms of the number of sensors or their locations with emphasis on two major criteria: the ability of sensors to capture gait patterns and the practicality of everyday life. This observation highlights the emergence of commercial wearable devices as a practical and user-friendly modality for gait monitoring in daily life.

Another trend which emerges from Table 3 is the fact of using several sensors together and this generally at various on-body locations [18, 64, 74, 85, 87, 91, 108, 145, 148, 154, 168, 173, 184, 205, 220, 256, 269, 279, 291, 303, 313, 314, 335]. However, using a multi-sensor system introduces several challenges, including in particular the integration of different sampling rates and signal amplitudes and how to align signals in multiple devices and therefore different clock times. Despite these challenges, the multi-sensor approach offers high potential for real-time monitoring of gait, where multi-sensor fusion can provide context-awareness (e.g., if the patient stays mainly at home or leaves it from time to time); and can contribute to the optimization of power (e.g., a low power sensor can trigger a higher power sensor only when necessary).

Another trend in ground-truth validation is increasingly in favor of using a reference (46 %) because of the confidence established from visually confirming the gait pattern being detected: this can be a metrological device (18 %), a medical examination (20 %) or a third-party portable medical device (8 %). However, in this case the data is not annotated and therefore does not allow conventional ML approaches.

Machine learning. The combination of ML algorithms and wearable sensors for gait analysis has shown promising results in validating the extraction of complex gait patterns [5, 8, 9, 21, 59, 76, 108, 145, 158, 159, 168, 191, 205, 208, 212, 252, 291, 303, 315].

As seen in Table 5, researchers have used ML on sensor data for different tasks : regression for continuous labelled data (speed, step length or distance) [76, 159, 205, 252] and classification of discrete labelled data such as groups of patients [8, 9, 21, 59, 158, 208, 252, 303] or or medical functional scores [5, 59, 145, 291, 315]. Classification, less commonly used for the validation of sensors, aims at higher level analysis, namely to identify a robust methodology able to monitor patients in time while at the same time discriminating between a pathological and physiological gait or the evolution of the disease studied on the basis of gait movements.

Type of ML algorithm families has evolved over time, with standard approaches before 2017 and the appearance for the first time in 2018 [191] of deep learning approaches with automatic feature extraction without human intervention, unlike most traditional ML algorithms. It should be noted that in the context of the papers studied in this review [108, 168, 191, 208], these approaches concern studies with a significant number of patients (> = 30) or/and relatively long acquisition times [191, 208] in order to guarantee a sufficiently representative and realistic sample. The other studies which are based on ML, because of samples often more limited in number of patients [21, 145, 158, 159, 205, 212, 252, 291, 315] or in acquisition time [5, 8, 9, 59, 303], preferred more standard approaches with a small number of expert features. Comparing the results of the different studies, in terms of performance, seems at this stage to be a difficult task because, as stated previously, it depends on the complexity of the task to be performed and the complexity of the ML algorithm implemented.

Finally, it should be mentioned that ML also has drawbacks, the first being the computational time required to train a model [235]. This is justified for complex analysis tasks such as classification or significant performance increase for a regression task. Moreover, ML may require the adjustment of hyperparameters which may demand theoretical knowledge in optimization. Finally, ML tends to be more difficult to interpret for a clinician who is looking for the most relevant parameters to analyze gait patterns of patients. However, it should be noted that recent initiatives have been carried out to demystify these two points [194, 312].

II.2.7.2 Recommendations

Advanced inertial sensors, including accelerometers and gyroscopes, are commonly integrated into smartphones and smart devices nowadays. So it is very convenient and cheap to collect the inertial gait data to achieve gait monitoring with high accuracy. Most existing validation methods ask the person to walk along a specified road (e.g., a straight lounge) and/or at a normal speed. Obviously, such strict requirements heavily limit its wide application, which motivates us to give here some recommendations for future work in this context.

Data collection and processing. A first step would be to define precisely validation protocols - by consulting the medical staff - adapted to the study of chronic pathologies. Indeed,

many studies only validate sensors for a given medical application without having tested them outside the laboratory, on a very limited number of patients, and over a relatively short time window (at most a few hours). The protocol to be defined should therefore impose experimentation constraints closer to the daily life of patients, namely: the data should be acquired at home, but also on a sufficient number of patients, over a sufficiently long acquisition period (several weeks, even months). It is also important to define which types of sensors would be more suitable according to the studied pathology, how many sensors would be necessary and where to place them on the patient because there is a clear trade-off between the accuracy of the recorded data and the invasiveness of the portable system.

It is mandatory to ensure that sensor recordings are accurate and sensitive enough for medical diagnosis and prognosis. This is crucial not only to ensure the generalizability of a sensor within a target population, but also its ability to measure day-to-day variability data which can be corroborated with disease symptoms. To this end, data acquired by commercial wearable sensors should be systematically compared to data acquired by reference medical devices (*i.e.*, reliable gold standard systems, medical scores or groups of subjects). ML approaches make it possible to loosen the strict framework of acquisition protocols but must ensure to collect large, labelled and realistic datasets for training. Deep approaches, which automatically select features from data, offer very interesting perspectives given that feature extraction is a task that can take teams of data scientists years to accomplish. It augments the powers of small expert teams, which by their nature do not scale.

Statistical models versus ML. Statistical models are designed for inference about the relationships between variables within the data and are designed for data with a few dozen input variables and small sample sizes. On the other hand, ML models are designed to make the most accurate predictions possible. Statistical models can make predictions, but predictive accuracy is not their strength. Indeed, no training and test set are necessary. Furthermore, ML aims at building a model that can make repeatable predictions in a high-dimensional space without formulating a hypothesis on the underlying data generation mechanism. ML methods are particularly useful when the number of input variables exceeds the number of samples [50]. Hence, using ML in a validation task highly depends on the purpose of the study. To prove that a sensor is able to respond to a certain kind of stimuli (such as a walking speed), then a statistical model should be used. Conversely, to predict from a collection of different sensors whether a patient is affected by a certain grade of a disease affecting the musculoskeletal system, ML is probably the best approach. Indeed, this multi-dimensional space (one or more for each sensor) is in fact difficult to interpret and therefore to analyze.

These are in fact data-based approaches which, as long as the data collected are numerous, annotated and representative, allow the training of an effective model. It should be noted that commercial wearable sensors allowing for increased data collection and good patient adherence through efforts of miniaturization, energy consumption and comfort will participate in this future success.

II.3 Security and privacy issues in ML

Although gait monitoring in healthcare is becoming more widespread over the past few years, these data are exposed to several security and privacy issues. We cover in this section a wide but non-exhaustive range of attacks and defenses related to sensitive information leakage from motion sensor data. This discussion allows us to place in a specific context the contributions in the next sections.

II.3.1 Threat model

It is necessary to define the environment, the actors, and the assets attacked in order to understand the ML attacks and defences. In general, the assets attacked concern the sensitive information which are the training dataset or the model itself with his parameters, hyperparameters and architecture. Four different entities are in most cases identified in practice, the data owners which correspond to the patient in rehabilitation application, the model owners which can be identical to the data owners and can share different information about the models or not. The model consumers, usually the hospital in case of rehabilitation application, use the model as a service. The last entity is the adversary try to access to different information depending on the application.

The adversary has access to a range of information which varies from the most limited without any auxiliary knowledge about the model such as the parameters and architecture or training data information, which corresponds to black-box attacks. The adversary can only provide inputs and observe the outputs provided by the model. On the other hand, the white-box attacks assume that the adversary has a complete access to the model parameters, hyperparameters and architecture or an access to the full dataset itself. A centralized learning architecture is concerned either by white-box and black-box attacks and FL architecture is mainly concerned by white-box attack either when the adversary is the centralized server or one of the participants.

An attack can be either passive or active. A passive attack occurs when the attackers are *honest-but-curious* which means that they do not interfere in the framework or the training procedure of the ML model, they only perform inferences by observing the system. In case of active attack, attackers actively modify part of the training process for example by being a malicious user that sends poisoned data to modify model training and maximise the misclassification error.

II.3.2 Sensitive inferences on motion sensor data

The main privacy threat concerns the access to sensor data to infer sensitive information. Some sensors such as GPS, cameras or microphones are well-known to be highly sensitive in terms of privacy [139, 170], it generally requires explicit permission from the user in order to be used by applications in smartphones for example [32]. Motion sensors, unlike the others, are still less known for their impact on privacy and also less protected [42, 336]. It is still possible to read in the literature some papers that ignore the privacy threats concerning motion sensors, for example by describing them as *"not particularly sensitive"* [327], or even consider the use of accelerometer as a privacy preserving method : *"The accelerometer-based approach does not require capturing privacy sensitive information"* [204]. The study of Crager and al. [66] highlights the fact that unlike traditional privacy issues associated with computers such as location or cameras, most users are not aware of novel privacy issues linked with mobile devices and sensors. We therefore present here several inference applications made on motion sensors data recording.

Behavior tracking We already saw in the section I.1 different ways to use motion sensors to track activities of patients, the number of steps, the distance walked or the energy expenditure. However, many other inferences less related to a rehabilitation program may be done to disclose potential sensitive information that the patient may not consent to give. Mannini and al. [201] showed that it is possible to detect more complex activities than walking or running, such as painting, writing, reading or sorting paperwork, and also if a person is carrying a load thanks to accelerometers placed on the wrist or ankle. Williamson and al. [330] estimate from a body-worn accelerometer the weight of carried objects in different configurations of walking condition and body types. Other researchers used devices worn on the wrist to detect even more specific activities such as eating and drinking moments [308, 350], smoking [270, 300]. Singh and al [286] used motion sensors worn inside a car to measure the behavior of a driver and identify events while driving such as sudden acceleration, breaking. Vaiana and al. [316] were able to identify aggressive and unsafe ways of driving, Dai and al. [69] predict if a person is drunk while driving. Accelerometer can also be used for voice detection and specifically detecting hotwords [349], indeed an acoustic signal can strike the inertial mass of the accelerometer that is sensitive enough to detect small changes in acceleration.

User re-identification The use of motion sensors at different locations on the patient's body to record for example hand gestures [292], head movements [188] or the gait directly with a smartphone on the waist or in the pocket [178], is also useful to uniquely recognize user from others. The methods previously described that use acoustic vibrations thanks to accelerometer for hotwords detection, have also shown that capturing such data allows the re-identification of different speakers with great precision [86]. Device fingerprinting is an approach that uses characteristics and features of devices to distinguish the users. Sanorita and al. [75] show that the hardware imperfections are made during the manufacturing of the sensors so that every sensor has his own response to the same motion stimulus with very subtle differences. Those differences are too small to impact higher functionalities but it allows researchers to discriminate different sensors with high accuracy. Those small anomalies in the signal response can be used to track users on the Internet and specifically the different websites visited even in private browsing mode or with cookies blocked [70].

Location and traffic patterns The leakage of location information on a user is a major issue as it can lead to the leakage of many sensitive information such as re-identification, social habits, home and work location, etc. That's why operating systems for smartphones developed mechanisms for users to manage control access to location services relying on Global Positioning Systems (GPS) or Wi-Fi. However, accelerometers, gyroscopes and magnetometers sensors on a smartphone can be used to infer vehicular users' location, traveled routes and the starting point by associating the sensor data with a map truth [130] or a graph generated from public database roads [223]. Han and al. [130] demonstrate that results obtained with motion sensors are similar to those obtained with a GPS, with an accuracy of 200 meter radius of the true location. Hua and al. [143] also reveals the device's location in the metro by only using the accelerometer sensor of smartphones. Results show an inference accuracy that reaches 94% if the user takes the metro for 6 stations.

Keystroke inference The inference of information concerning what a user is writing on his smartphone is obviously a highly sensitive threat as it concerns login credentials, personal notes, text messages. Motion sensors in smartphones such as gyroscopes, are able to detect

vibrations when typing on the screen and precisely the location of the vibration [51]. Features such as striking force of the fingers when typing, the resistance force of the supporting hand or the location of the supporting hand has influences on the shift and rotation of the smartphone captured by the gyroscope. Similarly to these methods, Aviv and al. [30] used accelerometers in smartphones to infer PINs and graphical password patterns. In the same way, Owusu and al. [236] could infer entire sequences of text.

Demographics parameters The estimation of demographic parameters such as age or gender has already been done thanks to descriptive statistic analysis: for example Menz and al. [211] discriminate young (22-39 years) and elder (75-85 years) users based on significant differences of step length, velocity and step timing variability. With ML, Davarci and al. [72] aimed at distinguishing child and adult users based on how they hold their hand and touch the smartphone while they are using an application. The inference of gender based on specific movement patterns has also been well studied with the extraction of hip movements [326], gait features [151] and different patterns of physical activity [150]. Weiss and al. [326] could estimate that gender recognition remains efficient regardless of the height and weight of the users.

Psychological traits inferences The emotional states of users has also been studied through the use of motion sensor data. Zhang and al. [352] could differentiate with high accuracy test subjects between three emotional states (happy, neutral and angry) based on accelerometer data recording on the wrist. Depressive mood has also been assessed [126], different stress levels [110] and arousal levels with smartphone carried on a pocket [234]. Some studies have also explored the inference of different personality traits such as extraversion, agreeableness, conscientiousness or neuroticism [26, 332].

II.3.3 Privacy risks on ML model

Recent research has shown that ML models are able to memorize sensitive information from training data and may occur unintentionally which means the sensitive information memorized is irrelevant for the learning task [53, 103, 109, 138, 287]. Artificial neural networks, in particular, have an enormous capacity to memorize arbitrary information [347].

When the adversary has no access to the data, he can attack the ML model itself during the training phase or the inference testing phase. Attacks on training aim at either learning information about the model or altering the model itself by injecting malicious training data. Attacks during the inference phase aims at inferring information by sending specific inputs to the model and observing the model's prediction without any knowledge about the model. Stealing a model can generally have two benefits, reuse the model for personal financial benefits, or to conduct other attacks that go further such as inferring information on the data itself. In the following paragraphs, we further describe some of these attacks on ML models.

Membership Inference Attack This adversary attack aims at inferring whether or not a specific data sample was used to train the targeting model or not. This information can raise privacy risks, for example it can identify the participation of a user to a health study training set in a hospital and then reveal that he was once a patient of this hospital. In the same way if the study is related to a specific disease, this attack may reveal that the patient has this disease. Membership inference attack can also measure the risk of potential privacy breaches that lead to other attacks. If the adversary, thanks to membership inference attack,

can infer that the victim's data is partly included in the information the adversary has access, then she can proceed to other attacks such as property inference [142].

The membership inference attack was firstly studied in a black-box context by Shokri and al. [285] in a supervised learning context. The attack exploits the fact that models generally behave differently to input data used for training and those never seen during training. It is commonly due to the overfitting effect of the model. Overfitting appears when the objective function fits a set of data points so closely that the model has better predictions on training inputs than inputs similar to the training population during the testing phase [272] (see Figure II.7). This results in higher confidence values (probability of the input class in the prediction vector) for training input than with data that does not belong to the training set. The method is further described with Figure II.8.

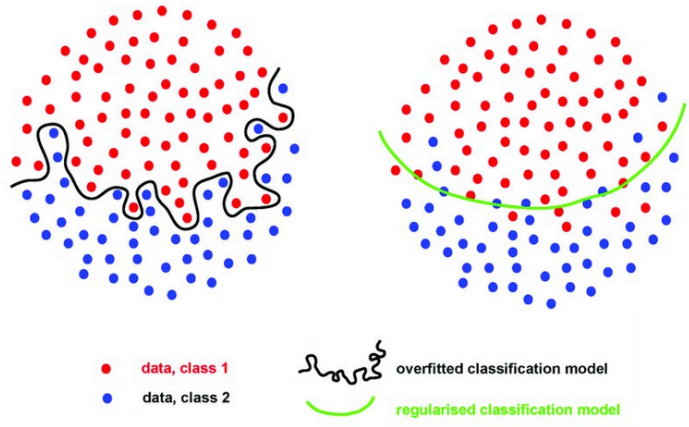


FIGURE II.7 – The concept of overfitting classification. In this case, the training error is much lower than test error. *Illustration reproduced from "From Big Data to Precision Medicine" Hulsen and al. 2019*

The goal is then to train a binary neural network called *attack model* in the Figure II.8, able to predict based on output's model if the input data belong to the training set D_{train}. To train the attack model, the adversary create k shadow models that reproduce the behavior of the target model. The adversary trains each shadow model with a dataset D'_k then produces a set of prediction P_k^m and P_k^{nm} based on the inputs D'_k and T_k a test set not used for the shadow model training. Once the output prediction sets are labeled, the target model can be trained to recognize whether or not an input data is a member of a training set, and finally test this model on target model output.

Salem and al. [271] propose a similar method than Shokri and al. [285] by relaxing two strong assumptions. Firstly, by showing that the attack does not need several shadow models and with only one the attack remains similarly efficient. Secondly, the dataset used to train the shadow models should no longer be from the same distribution as the target data through a data transfer method. These improvements tend to show that membership inference attack is widely applicable at a low cost.

Property inference attack consists of extracting dataset properties related to the users not explicitly correlated with the learning task, such as biometric information which can be learned by the model unintentionally. Even if the model is well trained, the property targeted may be relevant for the learning process and contribute to the high prediction performance of the model. This property can concern the entire dataset, for example an image classification model that aims at inferring the mood of a person (happy or sad) and can

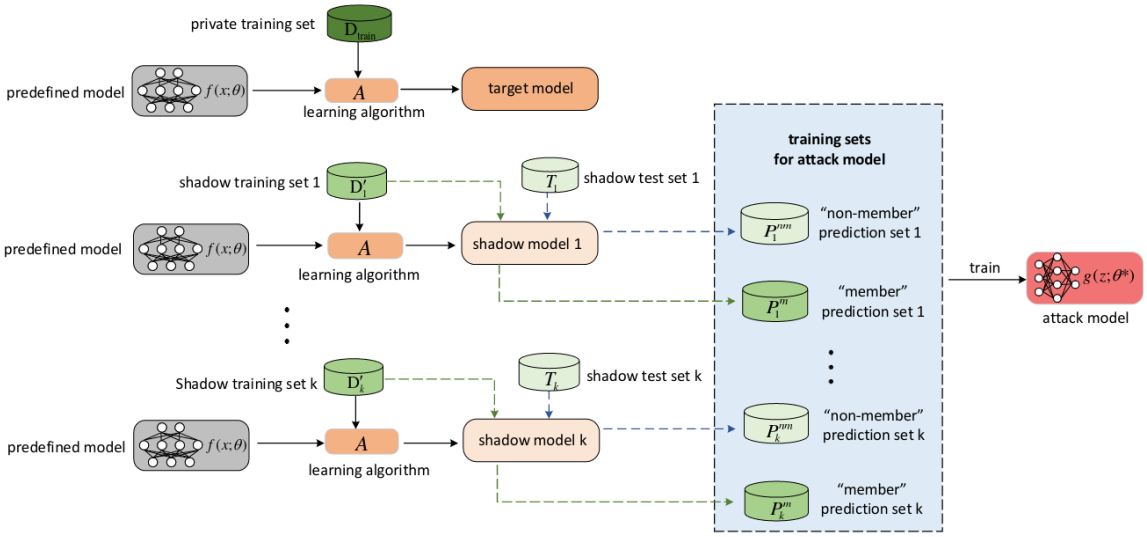


FIGURE II.8 – Membership inference attack based on shadow training. *Illustration reproduced from "Membership Inference Attacks on Machine Learning: A Survey" Hongsheng and al. 2021*

be used to infer the gender. The property targeted can also be inferred on a subset of the training dataset or even a specific user. Melis and al. [138, 209] performed an attack on a collaborative learning model as an adversary participant that tries to infer information about training data of a target participant. They especially focus on properties that are independent of the class characteristics. The authors take the following example: for an image classification that aims at inferring the gender, and for the subset of images that belongs to Bob, they infer that Alice appears in some photos.

Parisot and al. [239] propose a framework similar to membership inference attack with an attack model trained on shadow models specifically crafted to contain the target property or not. The training set is composed of the weights of the models of the shadow models. The attack model then infer whether or not the target model has the target property based on his weights.

In collaborative learning, this attack can be either passive or active. A passive adversary simply consists of observing the updates and performing inference without changing the training procedure. In the active way, the adversary trains his model to simultaneously infer the main task and the targeted property. By uploading these new updates for the aggregated model, this one learns a separable representation for the data with and without the targeted property [209].

This attack is further explored in Section IV.2 with an evaluation of the privacy issues of a specific FL configuration.

Model extraction attack This attack aims at stealing parameters of a model without prior knowledge about the model or the training data. Tramer and al.[310] was the first to present this attack by considering an adversary that can query a model in order to obtain the prediction vectors of input data . This attack can affect a wide variety of models such as decision tree, logistic regression, SVM or perceptron. Other work go further by reconstructing architectures and hyperparameters of deep neural networks [232, 321].

Model inversion attack Model inversion attack aims at partially or fully recreating training data with the associated class label thanks to a black-box or white-box access to the model. Fredikson and al. [103] introduced this concept on decision trees and neural networks applied on facial recognition. Their method uses prediction outputs of the model and estimates the probability for each possible input data to be the correct one that corresponds to the output. Other work, instead of reconstructing the training data, create class representatives or sensitive features that do not directly corresponds to the training data [104, 137, 339].

Evasion and Poisoning attacks Here, we focus on adversary attacks that aim at corrupting a target model and tampering his output. Two well-known types of attack are Evasion and Poisoning attacks.

An evasion attack aims at modifying the expected behaviour of a model by changing the input data [144]. This setting does not assume any influence over the training data. It can be done by adding an imperceptible noise to a normal testing sample such that the model gives incorrect predictions. A Poisoning attack alter the training dataset by inserting, modifying or deleting points in order to modify the decision boundaries of the targeted model [35, 165] and deteriorate the classification performance. The second possible objective is, as with evasion attacks, to modify the behaviour of the model in the direction that interests the adversary. In this latter case the Poisoning of the model preserves performance with one exception (a back-door), which has been chosen by the adversary [268].

II.4 Privacy-preserving ML schemes

This section introduces different common privacy preservation models. We focus on defence mechanisms directly related to the different contributions of the thesis: the contribution of Section III focus on data anonymization, Section IV.1 on data sanitization through adversarial networks methods and Section IV.2 explore the concept of differential privacy. Those aspects of privacy defences are then further detailed here.

II.4.1 Anonymization

Anonymization is the process of removing personal identifiers that lead to user identification. In practice, there is no universal anonymization solution that would apply to all types of applications and data. An anonymization solution must therefore be developed on a case-by-case basis and adapted to the intended use and the data to be processed. Thus, in order to demonstrate that a solution is correct and compliant with the GDPR, it must be shown that the anonymised data no longer allows the isolation of data belonging to an individual, no longer allows the linking together of distinct data sets relating to the same individual, nor does it allow the inference of information about an individual.

The anonymization techniques started in the late 1990s with several publications about the idea of k-anonymity model [299], which aims to prevent the unique identification of individuals from a small subset of their attributes, called a *quasi-identifier*. The subset of attributes to protect, which is not part of the quasi-identifier, represents the *sensitive attributes*. For example, within medical records, the birth date, sex and zip code triplet form a quasi-identifier that is enough to uniquely identify some individuals, while the disease is a sensitive attribute. k-anonymity states that to be protected, a user must not be distinguishable among at least $k-1$ other users. To do that, all k indistinguishable users must have the same values for all attributes forming their quasi-identifier. This makes them look similar and forms an *anonymity group*. Therefore, the probability of an adversary without external knowledge to re-identify someone among k similar users is at most $1/k$.

We denote a dataset d composed of points $X = \{x_1, ..., x_n\}$ and $A = \{a_1, ..., a_n\}$ features. We consider a quasi-identifier $Q_d = \{a_i, ..., a_j\} \subseteq \{a_1, ..., a_n\}$ associated with d and $r[Q_d]$ the projection of record $r \in d$ on Q_d. d respects k-anonymity if and only if each unique sequence of values in the quasi-identifier appears with at least k occurrences in d, or formally [248]:

$$\forall s \in \{r[Q_d], r \in d\}, |\{i \in \mathbb{N} | d_i[Q_d] = s\}| \geq k.$$

k-anonymous data is however vulnerable to several attacks. For example with background knowledge attacks, the combination of external information with a k-anonymous dataset can lead to privacy issues. With homogeneity attack, if each member of a given combination of identity-revealing traits has the same sensitive value, that value is revealed. These weaknesses of k-anonymity have been addressed by the introduction of ℓ- diversity [196]. It extends k-anonymity by ensuring a particular distribution of values for sensitive attributes across each anonymity group. The simplest way to do so is called distinct ℓ-diversity and states that there must be at least ℓ distinct values for each sensitive field for each anonymity group. t-closeness proposed by Liu et al. [186] is a further extension of ℓ-diversity. Instead of just guaranteeing a good representation of sensitive values, this approach enforces that the distribution of every sensitive attribute inside anonymity groups must be the same as the distribution of this attribute in the whole dataset, modulo a threshold t. The question of the usefulness of the data then begins to arise. Under the constraint of t-closeness, the data do not necessarily appear to be directly usable. However, it is still

possible to identify trends, or to perform general calculations or correlations on the whole dataset [187].

II.4.2 Differential privacy

Differential privacy firstly introduced by Dwork [81] is not an anonymization model, but the characteristic of an operation (or execution of an algorithm) on data that has certain guarantees of confidentiality. The idea is that an observer seeing the output of a differentially private algorithm is not able to tell if a particular individual's information was used in the computation. In other words, the addition or removal of one single element shall not significantly change the probability of any outcome of the aggregate function. Unlike k-anonymity, the differential privacy definition is not affected by the external knowledge an attacker may have. To formally define the differential privacy, we denote the privacy parameter $\epsilon \geq 0$, $X_{Q,d}$ a randomized output of a query Q on a dataset d and E_Q the range of Q. ϵ-differential privacy is satisfied when for all dataset d and d' differing from one element and for all measurable subsets $A \subseteq E_Q$:

$$Pr(X_{Q,d} \in A) \leq e^{\epsilon} Pr(X_{Q,d'} \in A).$$

One method to practically achieve differential privacy using numerical values relies on adding random noise following a Laplace distribution, whose magnitude depends on the *sensitivity* of the query function issued on the dataset. Intuitively, the sensitivity of a query function quantifies the impact that the addition or removal of a single element of a dataset could have on the output of this function [323].

Differential privacy has generated an important literature these last few years with new models and inter-model connections [78], as well as new techniques such as randomised response [325] and its combination with sampling [177] which achieves zero-knowledge privacy [112] (a privacy bound tighter than differential privacy).

II.4.3 Homomorphic encryption (HE)

Encryption process consists of encoding information by converting the original representation (called *plaintext*) into an unreadable format (called *ciphertext*) using a key, a piece of information when processed in the encryption algorithm can encode or decode data. In public-key cryptosystem the cryptography is asymmetric, the encryption key is public and distinct from the decryption key which is kept private. HE can be considered as an extension of a public-key cryptosystem and has the specificity to allow users of encrypted data to perform computations without decrypting the data. This scheme is quite suitable for cloud computing because a user can store encrypted data on an untrusted server that can perform computations without learning anything about the original data. A fully homomorphic encryption) has the following properties. We consider ciphertexts c_i and the corresponding plaintexts m_i. The decryption $Decrypt$ is described by those two properties:

$$Decrypt(c_1 + c_2) = c_1 + c_2$$

$$Decrypt(c_1 c_2) = c_1 c_2$$

which means that the decryption is doubly homomorphic, with respect to the addition and multiplication. This properties can be generalized with any function f a finite composition of addition and multiplication:

$$Decrypt(f(c_1, ..., c_t)) = f(m_1, ..., m_t)$$

FHE was then used with ML methods to maintain confidentiality [27, 122]. Dowlin et al. [117] was the first to propose a neural network based on encrypted data called CryptoNets. Several other initiatives start to leverage HE for ML [113, 135, 136]. For instance, [44] describes a method that consists of building blocks of homomorphic functions which they later use to compose several ML schemes. Although recent advances in HE schemes improve the performances [61, 79], these solutions are still resource consuming and face to scalability problems. In terms of available libraries, TFHE [304] provides the best performance for binary encoding while HEEAN [140] is the best one for floating point operation support.

II.4.4 Secure Multi-Party Computation (SMPC)

SMPC is a generic protocol used for distinct devices (or parties) to carry out a joint computation of functions knowing that each device wants to keep his input private. Yao and al. [340] was the first to propose a SMPC with 2 parties and Goldreich and al. [119] generalize it for a multiparty case. A SMPC protocol must respect several requirements:

— The *privacy* in the sense that devices needs to learn their output and nothing more

— The *correctness* means that each device is guaranteed that the output received is correct

— The *independence of input* means that the input of a corrupted device must be independent of the inputs of the other honest devices

— The *guarantee of output* means that a corrupted device is not able to prevent the other honest devices from receiving their outputs.

— The *fairness* guarantees that a corrupted device can receive his own output if and only if the other honest devices receive their outputs.

In an ideal world, the devices securely compute functions by sending their inputs to a completely trusted party that returns the output to all devices. In this configuration, an adversary can attack and take control of any devices but not the trusted party. This ideal world can be used as a benchmark reference to judge the security of a protocol. But in reality there is no trusted party. Instead, all devices communicate with each other with a specific protocol generally based on encryption methods [114]. ML as a service is an appropriate use case of SMPC as it would allow companies to propose their models to make inferences on the private data sent by their customers, while ensuring a privacy protection [273].

II.4.5 ML-specific approaches

ML techniques are also used to reduce information available to the adversary to mount their attacks. For instance, dropout is a famous technique often used to mitigate overfitting in neural networks. It consists of randomly setting neurons that make up hidden layers to 0 at each update of the training phase. This technique can also be used to reduce the effectiveness of Membership Inference Attacks based on overfitting [88].

Other techniques use GAN (see Section II.1) and the notion of adversarial networks competing in a mini-max game to build privacy-protection mechanisms. The objective of such methods is to learn in a competitive manner several neural networks that present different objectives. The auto-encoder mainly aims at generating, transforming or sanitizing the input data into a novel representation. This transformation is guided by the objectives of the other neural networks, which aims at preventing inferences on a specified sensitive attribute while still preserving the useful information contained in the data.

Several researches focus on preserving the task of activity recognition on motion sensors data. [199] is the only one which focuses on gender as the sensitive information. [197] in

their case, focus on the re-identification only while [253] apply their approach on several applications like object recognition or action recognition with several data types such as images or motion sensors. For these adversarial approaches that use autoencoders, the sensitive information can be extracted from the representation produced by the encoder [192], the decoder [253], or both the encoder and the decoder [197] for data sanitization. Specific to the sensor data generation, SenseGen [16] is a deep learning architecture for protecting users privacy by generating synthetic sensor data.

To enlarge with other applications protecting sensitive information using adversarial methods, [57] use a VGAN to transform face images in order to hide facial expression of the users that can be used to reveal their identity while preserving generic expressions. Adversarial approaches can also be used to hide sensitive information such as text in images [84] or identity information in the fingerprints [233].

From a broader privacy perspective, [311] proposes an adversarial network technique to minimize the amount of mutual information between a sensitive attribute and useful data while bounding the amount of distortion introduced. They applied their solution on a synthetic and a computer vision dataset. Inspired from [311], authors in [260] have developed a method for learning an optimal privacy protection mechanism also inspired from GAN, which they have applied to location privacy. In [240], authors have proposed an approach called table-GAN, which aims at preserving privacy by generating synthetic data. By suppressing *one-to-one* relationship and limiting the quality of dataset reconstruction, re-identification attacks are rendered less efficient.

Apart from techniques using adversarial approach to transform data representation to overcome privacy issues, [210] proposes two privacy preserving mechanisms based on clustering algorithms called Hierarchical Agglomerative Clustering to compress amount of disclosed data so that the amount of sensitive information can be reduced. [345] in their case, develop a framework for images data made on wearable cameras that can protect sensitive information such as face, objects or locations thanks to a neural network that detects the sensitive objects which will then be blurred or deleted. Rather than focusing on re-identification, [55] investigate what data to share, in such a way that certain kinds of inferences cannot be done. They propose *ipShield* that obfuscates data according to the quantification of an adversary's knowledge regarding a sensitive inference.

To summarize, this chapter provides an overview of research dealing with the two aspects of my thesis. The gait monitoring in healthcare and the impact of ML in this domain, and the different privacy issues implied by the production of motion sensor data and the use of ML models. The following chapter is dedicated to the first contribution in anonymization of motion sensor data.

Chapter III

Anonymisation through data minimization approaches

The two contributions presented in this chapter focus on the anonymization issue of motion sensor data. Both of them assume the configuration presented in Figure III.1. A user records motion data through IoT devices managed by a smartphone. We assume that the user application and the smartphone on which the application is run are trusted. Then the data are sent to an application server where they are processed and analysed. We assume that the application server runs on public cloud platforms. We consider that this cloud platform is honest but curious [118]. This means that the application server behaves correctly when it comes to processing data received from clients. More precisely, this means that the data is stored correctly in the database, that no forged information can be injected in the database, and that the classifier model cannot be maliciously tampered. However, we assume that the adversary is able to collect part or the entire information stored in the database and may also try to re-identify the user. In this context, re-identification is associated with the ability of a ML model to determine how different one user data is from other users [134, 246]. Each framework presented in this chapter tends to apply transformation and minimization methods in the sense given by the GDPR by limiting the *"collection of personal information to what is directly relevant and necessary to accomplish a specified purpose"* (see Article 5(1)(c) of GDPR [1]).

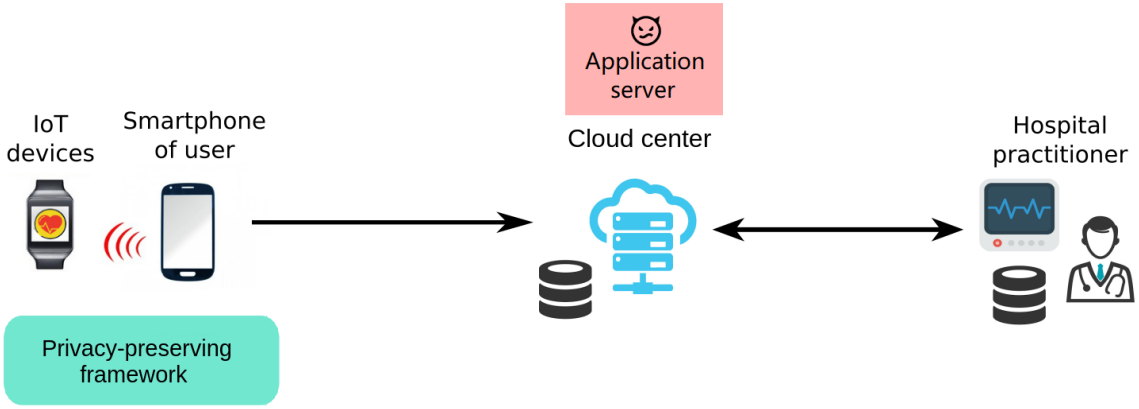


FIGURE III.1 – Illustration of a remote health monitoring system based on wearable sensors

III.1 Privacy framework for motion sensor data anonymization

In this section, we propose a privacy-preserving framework by firstly extracting well-known multiple features [284, 290] from raw signal and deeply analyse their impact on both the activity recognition and the user re-identification. We show that features in the

temporal domain are useful to discriminate user activity while features in frequency the domain lead to discriminating the user identity.

Based on this observation, we design a novel privacy-preserving framework. In this framework, data records are processed locally on the user device and only relevant features are extracted and sent to the server. Additionally, features in the frequency domain (*i.e.*, features leading to discriminate users) are normalized. This normalization can be viewed as a generalization-based approach. However compared to other generalization-based approaches based on *k*-anonymity that are well known to drastically reduce the utility of the protected data [123], our solution keeps a high utility (*i.e.*, activity recognition) while providing a good privacy (*i.e.*, small user re-identification). Once normalized, this information is periodically uploaded to the application server. Each batch of features is stored independently on the server (*i.e.*, with a different pseudonym) to avoid linking both batches to individuals and batches together.

Moreover, to avoid centralizing both the data and the associated identity of their owners on the same node, the mapping between the pseudonyms and the user identities is only retained by the hospital practitioners.

We exhaustively evaluated our ML framework with the use of two reference datasets. Results show an average accuracy of 87% in activity recognition while limiting the user re-identification rate up to an accuracy of 33% (for a dataset of 30 users). We also compared our solutions against different baselines. Our solution provides a better privacy-utility trade-off with a slight decrease of utility (9% drop in accuracy) against a large increase of privacy (53% drop in accuracy). In addition, by processing the signals at the edge of the network on the smartphone of users, our framework drastically reduces the operational costs on the application server (a decrease of 81% for computational cost). Lastly, we assess our framework with another dataset containing signals more perturbed by noise. In this case, we show that the impact on the accuracy of our framework remains limited and mostly impacts the static activities (e.g., standing activity). However, this impact can be removed by adapting the preprocessing step with filters according to the considered signals.

Our contributions can be summarized as follow:

— We quantify both the risk assessment associated with the re-identification of users (90% in average) and the capacity to detect the user activity (97% in average) from signals from mobile devices. Knowing that the state of the art in activity recognition is almost at the same accuracy [19]

— We deeply analysed the impact of multiple features on both the activity recognition and the user re-identification. We show that features in the temporal domain tend to discriminate the user activity while features from the frequency domain tend to discriminate users.

— We propose an efficient workflow and ML technique to recognise user activity with high utility while limiting the risk of user re-identification. Our solution provides a better privacy-utility trade-off with a slight decrease of utility (9%) against a large increase of privacy (53%) compared to state-of-the-art baselines, while reducing the computational cost on the application server.

— We test the capacity of our approach to be generalized by showing that the privacy-utility trade-off is better regardless of the considered classifier and also by assessing our framework with another smartphone dataset containing signals more perturbed by noise. We show a limited impact on the accuracy provided by our framework and we show that this impact can be removed by adapting the preprocessing according to the considered signals.

III.1.1 Methodology

This section explains the methodology we followed for activity recognition and user re-identification using IoT mobile devices. Although this description is specific to our methodology, it is typical and provides background on IoT healthcare workflow.

The whole workflow is depicted in Figure III.2 and includes data acquisition (Section III.1.1.1), signal preprocessing (Section III.1.1.2), segmentation (Section III.1.1.3), feature extraction (Section III.1.1.4), and classification (Section III.1.1.5). Figure III.2 also shows that the purpose (*i.e.*, the activity recognition) and one privacy risk (*i.e.*, user re-identification) are made through a common pipeline. These two tasks are done on the basis of classification with joint approaches (descriptors and ML algorithms). Section III.1.2 provides more details about the privacy risk assessment and the considered adversary model.

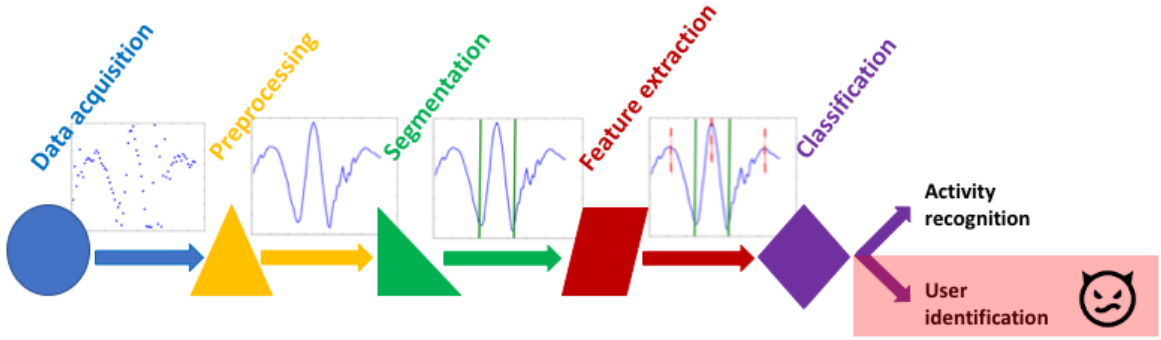


FIGURE III.2 – Traditional IoT healthcare workflow for activity recognition, an adversary can misuse the classifier to re-identify users.

III.1.1.1 Data acquisition

Data acquisition relies on sensors that are present in IoT devices, such as smartphones, smartwatches, smart wristbands, tablets and medical sensors. For the recognition of physical activities, we use here the inertial sensors accelerometer and gyroscope.

The data acquisition process is accomplished by a specific module in the mobile device and consists of the measurement and conversion of the electrical signals received by each sensor into a readable format [275]. Several challenges are associated with the data acquisition process when recognizing physical activities, including the positioning of the mobile device, the data sampling rate and the number of sensors to be used and hence managed [40]. All these factors directly influence the correct extraction of meaningful features. As the sensors are embedded in the mobile device, they cannot be located separately in different parts of the body; rather, the mobile device needs to be situated in a usual and comfortable position.

III.1.1.2 Signal preprocessing

Sensor signals are typically preprocessed by the application of a series of filters. First, noise was reduced with a median filter and a third order low-pass Butterworth filter with a cutoff frequency of 20 Hz. This frequency threshold was selected from the work presented in [163] which states that the energy spectrum of the human body motion is below 15 Hz. The resulting signals were further filtered to break them down into channels that make sense from a physical point of view as displayed in Figure III.3. For example, linear

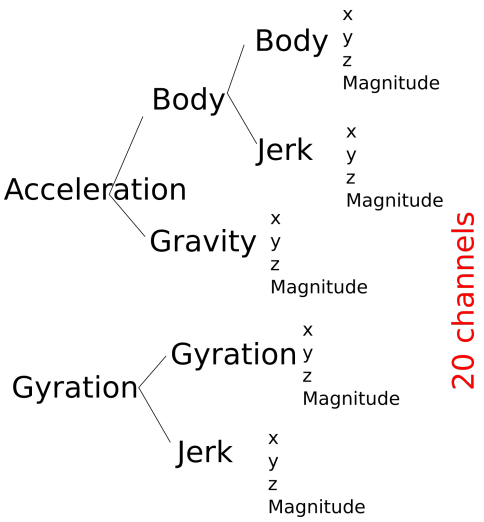


FIGURE III.3 – Channels considered for feature extraction.

acceleration signal was decomposed in two principal channels: gravitational and body motion components. This step was performed using another low-pass filter and assuming that the gravitational component mainly refer to the lowest frequencies [19]. Subsequently, body motion acceleration and gyration signals were derived in time to obtain jerk that reflect the temporal variations of the signals. Finally, signals were decomposed according to their acquisition axes (x, y, z, respectively) in order to observe them in a specific direction (vertical, lateral or longitudinal) as depicted Figure III.4. The magnitude of associated signals has also been calculated to produce an average signal less sensitive to how the device is fixed on the person. This filtering step allowed us to reach 20 channels in total.

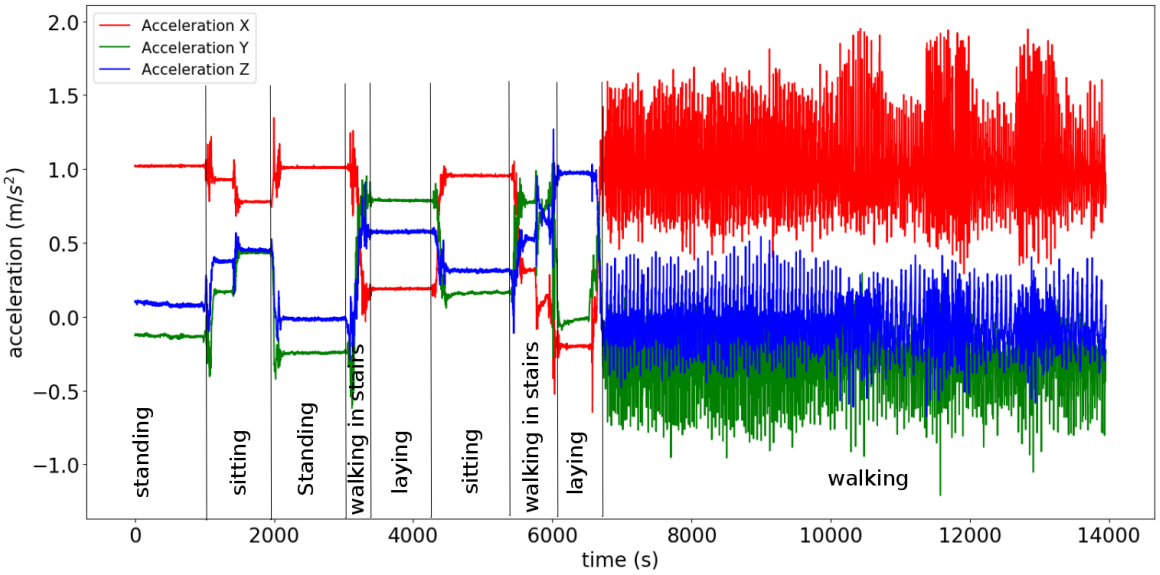


FIGURE III.4 – Visualization of accelerometer signals in x, y and z dimensions and associated activities.

III.1.1.3 Segmentation

Channel signals are typically segmented using a fixed sliding window technique. Windows with a span of 2.5 seconds and an overlap of 50% were captured. An overlap degree of 50% means that the window is shifted by half of its size, in other words 50% of the previous data are included in the next window. The choice of the window size is not trivial especially for an activity recognition algorithm. A small window size could split an activity signal while large window size could contain multiple activity signals. We decided to calibrate our window size on the most complex activity: walking. Hence, the window size has been chosen to take into account at least a full walking cycle of two steps: the cadence range of an average person walking corresponds to minimum speed of 1.5 steps by second according to [37].

	Function	Description	Formulation
Time domain	mean $(\boldsymbol{s})$	Arithmetic mean	$\bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i$
	std $(\boldsymbol{s})$	Standard deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(s_i - \bar{s})^2}$
	mad $(\boldsymbol{s})$	Median absolute deviation	$\text{median}_i\left(\lvert s_i - \text{median}_j(s_j)\rvert\right)$
	max $(\boldsymbol{s})$	Largest values in array	$\max_i(s_i)$
	min $(\boldsymbol{s})$	Smallest value in array	$\min_i(s_i)$
	sma $(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3)$	Signal magnitude area	$\frac{1}{3}\sum_{i=1}^{3}\sum_{j=1}^{N}\lvert s_{i,j}\rvert$
	iqr $(\boldsymbol{s})$	Interquartile range	$\text{Q3}(\boldsymbol{s}) - \text{Q1}(\boldsymbol{s})$
	autoregression $(\boldsymbol{s})$	4th order Burg Autoregression coefficients	$\boldsymbol{a} = \text{arburg}(\boldsymbol{s}, 4), \boldsymbol{a} \in \mathbb{R}^4$
	correlation $(\boldsymbol{s_1}, \boldsymbol{s_2})$	Pearson Correlation coefficient	$C_{1,2}/\sqrt{C_{1,1}C_{2,2}}, C = \text{cov}(\boldsymbol{s}_1, \boldsymbol{s}_2)$
	angle $(\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, v)$	Angle between triaxial signal mean and vector	$\tan^{-1}\left(\lVert[\bar{s}_1, \bar{s}_2, \bar{s}_3] \times \boldsymbol{v}\rVert, [\bar{s}_1, \bar{s}_2, \bar{s}_3] \cdot \boldsymbol{v}\right)$
Frequency domain	skewness $(\boldsymbol{s})$	Frequency signal Skewness	$\text{E}\left[\left(\frac{\boldsymbol{s}-\bar{s}}{\sigma}\right)^3\right]$
	kurtosis $(\boldsymbol{s})$	Frequency signal Kurtosis	$\frac{\text{E}\left[(\boldsymbol{s}-\bar{s})^4\right]}{\text{E}\left[(\boldsymbol{s}-\bar{s})^2\right]^2}$
	maxFreqInd $(\boldsymbol{s})$	Largest frequency component	$\arg\max_i(s_i)$
	energy $(\boldsymbol{s})$	Average sum of the squares	$\frac{1}{N}\sum_{i=1}^{N} s_i^2$
	entropy $(\boldsymbol{s})$	Signal Entropy	$\sum_{i=1}^{N}(c_i \log(c_i)),\ c_i = s_i / \sum_{j=1}^{N} s_j$
	meanFreq $(\boldsymbol{s})$	Frequency signal weighted average	$\sum_{i=1}^{N}(is_i) / \sum_{j=1}^{N} s_j$
	energyBand $(\boldsymbol{s}, a, b)$	Spectral energy of a frequency band $[a, b]$	$\frac{1}{a-b+1}\sum_{i=a}^{b} s_i^2$

FIGURE III.5 – List of measures for computing feature vectors. N: signal vector length, s: the signal vector, $s_{1,2,3}$: x, y an z vector of a signal, v: the base vector to measure the angle (y axis [0,1,0] selected), (a,b) the frequency band (three different bandwidth selected: 8, 16 and 24 points), Q: quartile.

Xacc_body_iqr	Xacc_body_max	Xacc_body_mean	Xacc_body_med	Xacc_body_min	Xacc_body_ropy	Xacc_body_std		pers	act
0.77666792327	1.01481659060	0.32071585656	0.34988767835	-0.49054102707	4.7489565343	0.4194744014		10	2
0.66693512370	1.43263647481	0.26841672908	0.43411692212	-1.41238613704	4.7722314297	0.6610481443		10	3
1.02907915173	1.43263647481	-0.10075092775	0.08232560553	-1.42686548654	4.7899910551	0.6334978541	...	10	3
0.23557396729	0.74911155782	0.33652443467	0.26582976888	0.10360618631	4.8056637254	0.1568877474		10	4
0.35584093169	0.78654658654	0.21654485464	0.27026656791	-0.72443435405	4.7194139887	0.3592030590		10	4

FIGURE III.6 – A sample dataset with features and labels, input of the classification step.

III.1.1.4 Feature Extraction

From each window of each channel signal, a feature vector was extracted which contained 17 measures estimated in the time and frequency domains respectively. The magnitude of the Discrete Fourier Transform (DFT) was used to extract the descriptors of each window in the frequency domain. The choice of these descriptors was made on the basis of an earlier review on effective descriptors for gait recognition [290] : e.g. for time domain mean, standard deviation (STD), signal magnitude area (SMA) and signal-pair correlation (Corr); and for frequency domain energy and entropy. The selected measures to obtain the feature vector are depicted in Figure III.5. A feature vector was calculated from each experiment window sample and labeled according to the user and activity it belongs. Figure III.6 shows an example of the dataset format, where lines correspond to window samples and columns to features (except the two last ones which correspond to the labels). Such dataset is used as an input for the classification task. A total of 510 features are extracted. The notation for naming a descriptor in the rest of this article is the following $\{orientation\}\_\{channel\}\_\{descriptor\}$. *orientation* represents the axis (x,y,z) or the magnitude (magn). *channel* represents the group of channel considered such as Body, Jerk, Gravity and Gyration. *descriptor* represents the feature considered such as Mean or Standard deviation (std).

III.1.1.5 Classification

ML algorithm There are multiple ML algorithms that can effectively handle these features (e.g., Decision Tree, Support Vector Machine, RF). We evaluated a number of them (one representative for each ML family) as illustrated in Table III.1. In order to make a fair comparison, the different algorithms were optimized independently. From this analysis, it follows that it is RF that provided the best results for our use case. Consequently, RF was chosen for the multi-class classification tasks in the remainder of this study.

In general, the RF algorithm is a supervised classifier having fast training time and very high performance without fine-tuning [207]. The function "RandomForestClassifier" of the Python Scikit Learn package [244] was used to build the RF classifier and related to its optimization, 700 was chosen as the number of trees in the forest, $\sqrt{n}$ random features were considered in building each tree, with n representing the number of features (using square root function force the trees to be limited in the depth in order to avoid overfitting), and 10 was set as the maximum depth of each tree.

Utility and privacy measures To measure the classification quality based on the proposed features with RF, we computed the accuracy from the confusion matrix [129]:

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FP| + |FN|},$$

where $|TP|$ (True Positive): is the number of correct predictions when the actual class value and the predicted class value are the same and correspond to the targeted class, $|TN|$ (True Negative): is the number of correct predictions when the actual class value and the predicted class value are different from the targeted class, $|FP|$ (False Positive): is the number of incorrect of predictions when the actual class value and the predicted class value are the same and correspond to the targeted class, and $|FN|$ (False Negative): is the number of incorrect predictions when the actual class value and the predicted class value are different from the targeted class.

Accuracy reflects the number of correct predictions made by the model over all kinds of predictions made. Accuracy is comprised in $[1:0]$ where a value of 1 corresponds to a perfect prediction. We prefer the accuracy rather than the f-score because the variable classes in the data are nearly balanced. We use this metric to compute the quality of our classification to predict both the activity of the user and the user identity. We leverage this metric to define a utility and a privacy measurement. Specifically, we called *Accuracy(activity)* the result of the accuracy when it is applied to the activity recognition (utility metric), and we call *Accuracy(re-identification)* the result when it is applied to the user identity (privacy metric).

Algorithme 1 : Feature selection

Input : List of features sorted by importance f and associated initial accuracy a; $threshC = 0.7$; $threshA = 0.03$

Output : List of selected features

```
1  for each feature f_i ∈ f do
2      Compute the Pearson correlation values C for each feature in {f − f_i} : fcorre
3      for each feature f_j ∈ fcorre do
4          if | C(f_j) |> threshC then
5              Compute accuracy newA of classification for {f − f_j} : newa
6              if a − newA < threshA then
7                  Erase feature f_j from f
8              end
9          end
10     end
11 end
```

Feature ranking and selection The RF algorithm can also be used to rank features according to their importance in the classification. When training a tree, it can be computed how much each feature decreases the Gini impurity index [152] in a tree. For a forest, the impurity decrease from each feature can be averaged and the features are ranked according to this measure.

The RF algorithm can also be used for feature selection [46]. This is done via measuring the mean decrease of accuracy when a particular feature is removed from the set of features in the trees. If the accuracy deterioration after feature exclusion is negligible, the feature is less important and vice versa. The importance scores of the features in the RF classifier [46, 124] can therefore be evaluated and used as a feature selection criteria. For more details, see the Algorithm 1: It consists of two nested loops, one corresponding to features ranked by importance (line 1) and one corresponding to features correlated to each of the features of the first loop (line 3). The correlation is calculated using the Pearson coefficient (line 2). If the correlation between two features is greater than a certain threshold (line 4), then the accuracy of the RF algorithm is recalculated after removal of the correlated feature (line 5) and if the corresponding decrease in accuracy is below a certain threshold (line 6) this feature is eliminated for good (line 7).

III.1.2 Adversary model

This section presents the architecture of a traditional centralized system without any protection and the potential attack we want to protect the system from (Section III.1.2.1). Then we present the assumptions made to design our solution (Section III.1.2.2).

III.1.2.1 Traditional architecture

In this traditional architecture, (❶) IoT devices or directly the smartphones perform the data collection from sensors and (❷) send the raw data to the application server which stores them. (❸) The server then performs all the remaining tasks including the preprocessing, the segmentation, feature extraction, and the classification of the activity. Finally, (❹) the hospital practitioner requests the application server to have an analysis of the activity of patients.

This centralization of the raw data exposes users to many privacy risks in case of data leak. Indeed, if the server is compromised or if some data are stolen, raw data are revealed leading to the possibility to do many sensitive inferences including re-identification. In this work, we focus our privacy assessment on this user re-identification risk.

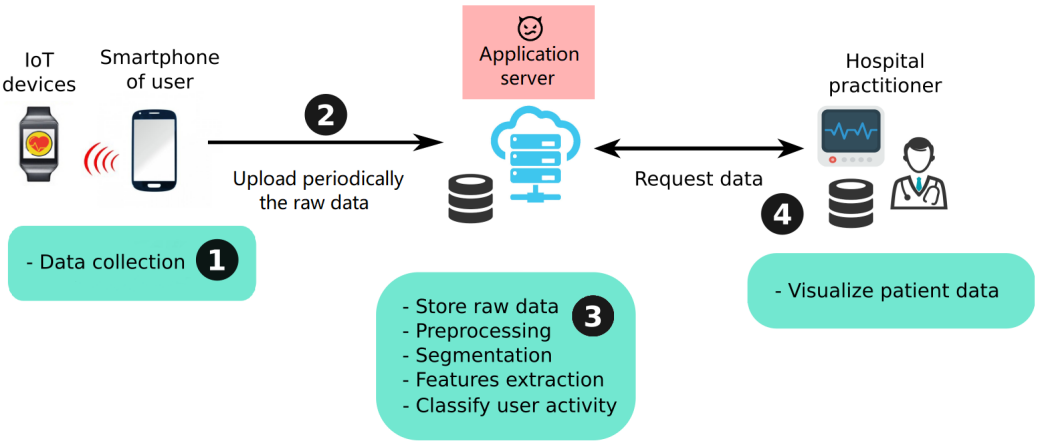


FIGURE III.7 – A traditional architecture: the user smartphone send directly the raw data to the application server that upload it periodically.

III.1.2.2 Technical assumptions

Before presenting our privacy-preserving framework in Section III.1.4 and after giving some common assumptions at the beginning of the chapter, we further describe our assumptions and the adversary model against which this solution is designed. The framework presented in this section involves three premises: the client running on the smartphone of users, the application server storing the features and performing the classification, and the hospital practitioner monitoring the patient activity. First, a trusted client application means that the data acquisition, the preprocessing, the segmentation, the feature extraction, and the normalization cannot deviate from a correct behaviour. Moreover, we do not consider limitations on the sampling rate of the data acquisition as in [301].

Second, each information stored on the application server corresponds to independent batches of data unlinked to users (*i.e.*, with a different random pseudonym for each batch). Additionally, we assume that the adversary is able to collect data relative to the gestures of each user from a malicious IoT device for instance. This prior knowledge on each user is used by the adversary to build a classifier model. This classifier exploits the same preprocessing, segmentation, and features than our classifier but with the objective to predict the identity of the user for each batch of data stored in the database.

Third, we assume that the server used by the hospital practitioner is trusted. This server is used to store the mapping between the batches of data sent to the application server and the identity of the users.

Lastly, all communications between nodes (*i.e.*, clients, the application server, and server of the hospital practitioner) are secured. We assume that no information can be inferred from these secured communications.

III.1.3 Quantifying activity recognition and user re-identification

We carried out an extensive evaluation of the capacity to recognise the activity of users and to re-identify them. We show that following the methodology described in Section III.1.1, we are able to predict the activity of the user with a very high rate of success. In addition, we show that without any protection scheme, data from mobile devices act as a personal fingerprint and lead to re-identify users. We first describe the dataset used in this evaluation (Section III.1.3.1) before to quantify the activity recognition and the user re-identification (Section III.1.3.2). Finally, we analyse the impact of extracted features (Section III.1.3.3).

III.1.3.1 Dataset

The dataset used in this work is available online for public use as the "Human Activity Recognition using Smartphones" dataset in the UCI Machine Learning Repository [19]. This dataset represents a reference for evaluating activity recognition learning models. It is composed of the 3-axial raw data from accelerometer and gyroscope sensors read at a constant frequency of 50 Hz. A group of 30 volunteers were selected to follow a protocol of activities while wearing a smartphone on their waist. The experiment was planned in order to contain six basic activities: three static postures (standing, sitting, lying-down) and three ambulation activities (walking, walking-downstairs and walking-upstairs). Figure III.4 displays accelerometer signal of one of the experiments and the associated activities. The protocol of activities is detailed in [257]. The duration of an entire experiment was around 15 minutes and was repeated ten times. All the experiments were recorded on video to have a ground truth to annotate the performed activities on acceleration and gyration signals. The data of each user is splitted into train, validation and test subset in order to have all the users in each subset. The data firstly splitted into 80% and 20% for train/validation and test subsets. Then splitted into 80% and 20% for train and validation.

Algorithm	Accuracy (activity)	Accuracy (identity)
Decision Tree	0.94	0.73
K-nearest Neighbors	0.78	0.36
Support Vector Machine	0.58	0.23
Gaussian Naive Bayes	0.80	0.14
Random Forest	0.96	0.82
Quadratic Discriminant Analysis	0.88	0.63

TABLE III.1 – Comparison of different well-known algorithms in terms of activity and identity performance.

III.1.3.2 Activity Recognition and User Re-Identification

We firstly evaluated the accuracy of different well-known classification schemes for the activity recognition and the user re-identification in order to select the best one for our use case (Table III.1). Without optimizing parameters (*i.e.*, using standard values), RF outperforms other schemes for both classification tasks with 0.96 and 0.82 of accuracy for activity recognition and user re-identification, respectively. Once the most adapted classification

Activity	Accuracy(activity)
Walking	0.97
Walking upstairs	0.95
Walking downstairs	0.94
Sitting	0.97
Standing	0.98
Laying	0.99

TABLE III.2 – User activities can be recognised with a high success rate (recognition using the methodology presented Section III.1.1).

scheme identified, we then optimized the following parameters to further increase the accuracy : the number of trees in the forest, the maximum number of features considered for splitting a node, the maximum number of levels in each decision tree. For the optimization method we use *GridSearchCV*, an *sklearn* method that generates all the combinations of hyperparameter following the grid of possible hyperparameter given and provide the optimal set of hyperparameter according to the accuracy function given.

Table III.2 summarizes the accuracy for the recognition of the different activities. Results show that our ML framework is able to highly recognise activities with an average accuracy of 0.97 which is comparable to state-of-the-art performance [164]. As the table indicates, the accuracy is lower for ambulatory activities in stairs. A possible explanation for this is that these activities correspond to the smallest acquisition times (Figure III.4).

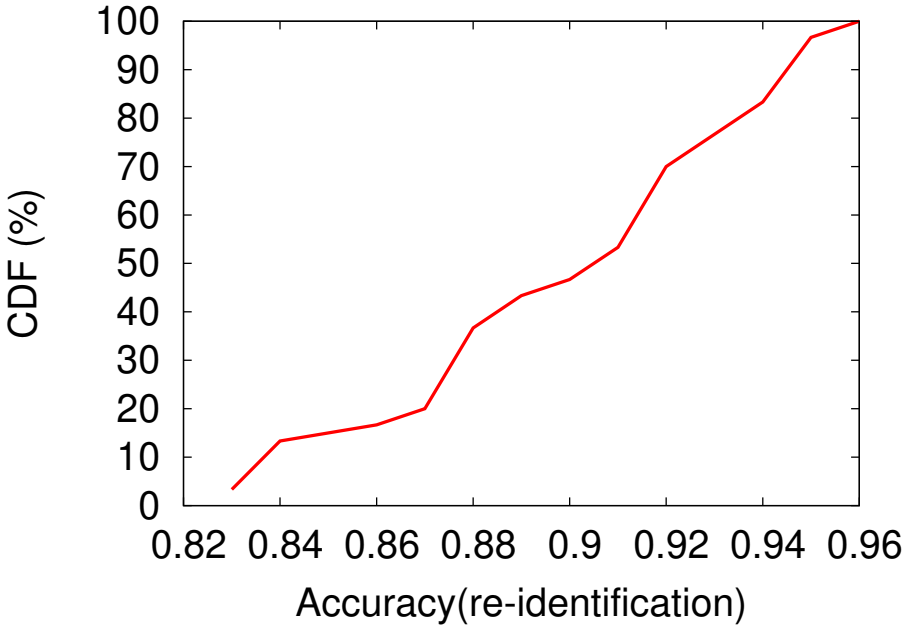
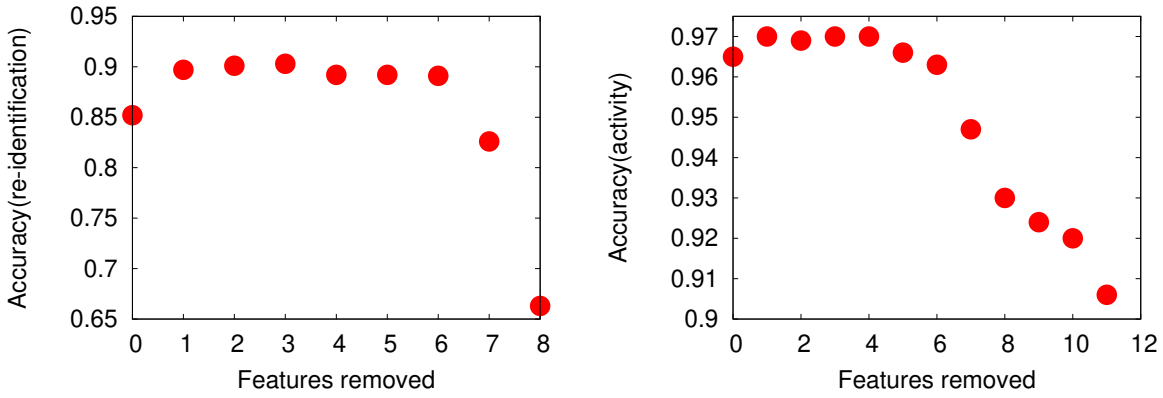


FIGURE III.8 – Cumulative distribution of the accuracy for the user re-identification task: users can be easily re-identified from their data.

Figure III.8 depicts the cumulative distribution of the accuracy for the user re-identification task. Accuracy ranges from 0.82 to 0.96 among the 30 users with an average of 0.90. These results indicate that the data collected from the gesture of users characterizes each individual and can lead to re-identify them with a high success rate. However, the task of



(A) Privacy (B) Utility

FIGURE III.9 – Impact of the number of features (depicted in Table III.3 and Table III.4) retained in the RF learning process on user's privacy and utility metric (features were sorted by increasing order of importance).

re-identification is slightly more difficult than that of recognizing activities with lower accuracy.

III.1.3.3 Impact of Features

The previous experiments are also used to rank features (from the 340) according to their importance. Eight and eleven features were respectively selected for the activity recognition and user re-identification tasks given the correlation and accuracy analysis (see Algorithm 1 for methodology and Tables III.3 and III.4 for results) with one temporal feature in common (Magn_grav_max). Indeed, many features are alike and contain similar information on the original sensor data. Compared to using all 340 features, using only these 19 relevant features lowers only slightly ($< 4\%$) the two classification tasks performance (97% vs 96% for activity classification and 90% vs 86% for user re-identification).

This can be observed more precisely in the Figures III.9a and III.9b, where the importance of each selected feature is independently tested for the task of interest: there is a strong correlation between the importance of a specific feature and the performance of the RF algorithm after removing it.

Based on these ranking results, it is interesting to note that the task of activities recognition (*i.e.*, utility) is almost exclusively (9 of the 11 selected features) operated in the time domain whereas the task of user identification (*i.e.*, privacy) is based (5 of the 8 selected features) on features in the frequency domain. These results can be explained by the fact that the activities are mainly distinguished from each other by their level of amplitude in acceleration and gyration (Figure III.4) and therefore their associated statistics. Conversely, the user identification is more related to the pace or cadence at which this person performs the activity and is strongly related to biomechanics (e.g., age, size, weight).

III.1.4 Privacy preserving activity recognition framework

To ensure privacy, our framework relies on both an architecture limiting the exposure of sensitive information and a data normalisation applied on features leading to re-identify

Features	Importance
Y_grav_std	0.175
Z_grav_med	0.163
Z_grav_energy	0.137
X_grav_max	0.128
Magn_grav_max	0.123
Y_gyro_mean	0.107
Y_gyro_irq	0.088
Y_body_zcross	0.079

TABLE III.3 – Most important features for user re-identification (frequency-based features are in grey).

Features	Importance
X_grav_max	0.144
X_grav_min	0.127
Magn_grav_max	0.109
X_gyro_min	0.104
X_body_var	0.098
Magn_body_var	0.085
X_gyro_max	0.082
Y_gyro_irq	0.078
X_gyro_mean	0.077
Magn_gyro_mean	0.074
Y_body_entropy	0.020

TABLE III.4 – Most important features for activity classification (frequency-based features are in grey).

users (Section III.1.3.3). These normalisations act as a form of generalisation-based obfuscation. In this section, we first present the architecture of our framework (Section III.1.4.1) before to describe the normalisation of each sensitive feature (Section III.1.4.2).

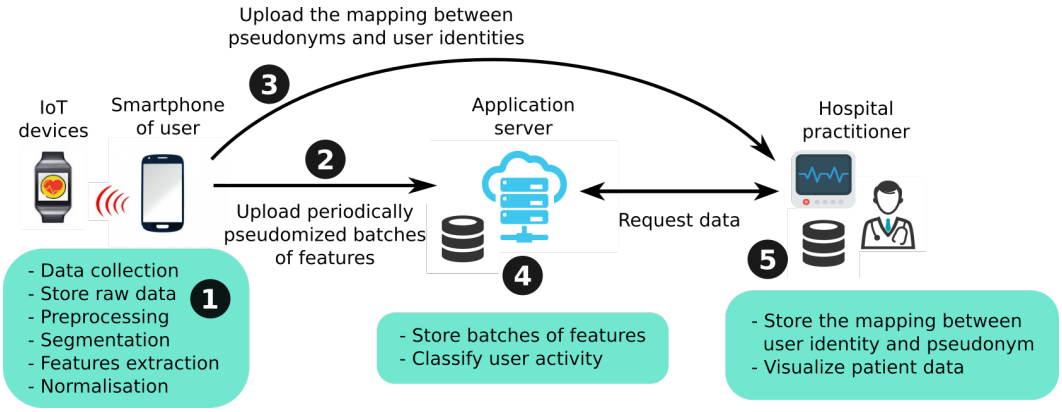


FIGURE III.10 – Architecture of our framework: the user smartphone is leveraged to extact relevant features and only these features are uploaded periodically to the application server.

III.1.4.1 Architecture

The design of our privacy-preserving framework comprises three main elements: a client application running on the user smartphone communicating with its IoT environment, the application server, and the hospital practitioner. To limit the exposition of sensitive information, the application server does not store identified data but only batches of features where each batch is randomly pseudo-anonymized. Only the hospital practitioner retains the mapping between the user identities and pseudonyms, and requests the application server to monitor the activity of users.

The architecture of our privacy-preserving activity recognition framework is depicted Figure III.10. Firstly, (❶) IoT devices (e.g. smartwatch) or directly the smartphones perform

the data acquisition. In both cases, these raw data are stored locally on the smartphone. The client application then performs the preprocessing, the segmentation and the features extraction following the methodology described in Section III.1.1. On the basis of our analysis on the importance of features, this feature extraction only concerns the 19 features identified as important (Section III.1.3.3). Moreover, the client conducts the normalisation of the features identified as leading to the re-identification of users. All these normalisations are described in the following sub-section. As all the aforesaid actions performed on the smartphone only concern the associated user on one batch of data (*i.e.*, one day for instance), the resulting computational cost is cheap (Section III.1.5.3).

Secondly, (❷) the client application associates a random pseudonym to each timestamped batch of features before to periodically upload them to the application server. (❸) The client application then sends to the hospital practitioner the list of pseudonyms associated to its identity.

(❹) When a batch of features is received by the application server, it stores this information in a database. Consequently, each batch in this database does not contain the identity of the user but a random pseudonym. The application server then periodically performs the classification to detect the activity associated with each batch of features.

Finally, when the hospital practitioner wants to monitor the activity of a specific user, firstly it retrieves locally all the pseudonyms associated with the specified user and then requests the application server to have the activity history of the specified pseudonyms (❺). In this approach, the requests from hospital practitioners could lead to link different pseudonyms to the same patient. To overcome this problem, fake requests could be sent to hide the ones targeting the expected patient.

III.1.4.2 Normalisation

In order to limit the re-identification of users, we propose a normalisation approach which generalises the effect of the different descriptors identified as important for the task of user re-identification. Similar to k-anonymity which ensures anonymity group gathering k different users with the same quasi-identifier, our normalisation approach aims at erasing the characteristics of a single specific user (*i.e.*, leading to the re-identification) and transforming the data so that, after normalisation, the data of all users share the same statistical characteristics.

Given the data from the sensors noted S and of size n, applying the normalisation approach on S will output the so-called "normalised data" noted S^*. In this work, we distinguished five normalizations, each of them referring to the features in the frequency domain listed in Table III.3 (the feature corresponding to the normalisation is given in parentheses).

For the three temporal features that remain in Table III.3, we proposed to delete the two of them that were not used for activity recognition. We kept the last one because it was also selected for activity recognition in Table III.4. S is a set of raw data and S_i an element of the set with n the length of the set. In our case the set corresponds to a data window. All the windows are normalized so that we can no longer discriminate the users with this information. All the normalization presented are linear operations so they can be combined.

Normalisation by mean (Y_gyro_mean)

$$S_i^* = S_i - \mu + \mu^*, \quad i \in [0, n], \quad \mu^* = 0, \tag{III.1}$$

with μ and μ^* being respectively the data means before and after normalization.

Normalisation by interquantile range (Y_gyro_irq)
The interquantile range (IQR) is a measure of statistical dispersion, being equal to the difference between 75th and 25th percentiles.

$$S_i^* = \frac{S_i}{IQR} IQR^*, \quad i \in [0, n], \quad IQR^* = 1, \tag{III.2}$$

with IQR and IQR^* being respectively the data interquartile ranges before and after normalisation.

Normalisation by standard deviation (Y_grav_std)

$$S_i^* = \frac{S_i}{\sigma} \sigma^*, \quad i \in [0, n], \quad \sigma^* = 1, \tag{III.3}$$

with σ and σ^* being the data standard deviations before and after normalisation.

Normalisation by root mean square (Z_grav_energy)

$$S_i^* = \frac{S_i}{\sqrt{\frac{1}{n}\sum_{j=1}^{n} S_j^2}}, \quad i \in [0, n]. \tag{III.4}$$

Normalisation by maximum and minimum (X_grav_max)

$$S_i^* = (S_i - Min)\frac{newMax - newMin}{Max - Min} + newMin, \quad i \in [0, n], \quad newMax = 20, \quad newMin = 0, \tag{III.5}$$

with Max and Min being respectively the maximum and minimum of the original data and $newMax$ and $newMin$ the maximum and minimum of the normalised data.

III.1.5 Evaluation of the framework

We carried out an extensive evaluation of our framework. In this section,
we start with a description of the comparison baselines (Section III.1.5.1) before evaluating the performance of our approach in terms of utility-privacy trade-off (Section III.1.5.2).

III.1.5.1 Comparison Baselines

To highlight the benefits of our approach, we compare the performance of our framework with that of three alternatives. The first alternative follows a perturbation scheme. Similarly to the differentially private approach described in [6] that applies a perturbation scheme in the frequency domain of aggregated time series in the context of location privacy, this alternative (called *perturbation*) adds a Gaussian noise in the signal in frequency domain before the extraction of features. The noise is added to the channels used for re-identification. The authors in [6] showed that with Gaussian random variable more concentrated around 0 than a Laplace random variable, perturbation scheme cound ensure better utility. The second alternative is based on simply the removal of features identified as

leading to the user re-identification (Section III.1.3.3). The incentive behind this alternative (called *suppression*) is that without these features, the re-identification is harder. The last alternative is a privacy-preserving classification based on homomorphic encryption. This alternative implements a RF classifier working over encrypted data similar to [44]. In this solution (called *homomorphic*) the input data (*i.e.*, the features used by the RF model used in Section III.1.3.2) are encrypted by the smartphone before to be sent to the server which is able to do the classification of the activity directly from these encrypted data. To achieve that, the multiple decision trees of the RF classifier are expressed as a polynomial P whose output is the result of the classification. More precisely, each node in the trees is a boolean variable defined at 1 if, on input x, one should follow the right branch, and 0 otherwise (*i.e.*, the value of a variable b_1 is 1 if the input x_1 is smaller than the threshold w_1, and 0 otherwise).

Consequently, P is a sum of terms, where each term t corresponds to a path in a tree from root to a leaf node c. A term t evaluates to c only if an input x is classified along that path in the tree, else it evaluates to zero. Hence, the term corresponding to a path in the tree is naturally the multiplication of the boolean variables on that path and the class at the leaf node. We use TFHE [304] to implement this solution in C++ and the value of the inputs as well as the threshold are coded in 16 bits.

III.1.5.2 Privacy Improvement

Figure III.11 reports for our solution and the baseline approaches the trade-off between the utility captured by the accuracy to recognise the activity and the privacy captured by the accuracy to re-identify users. For the baseline based on the suppression of features, each point of the curves corresponds to the deletion of a feature (from the 8 selected ones for the re-identification task) from the less important to the most important feature. For the baseline based on perturbation, in turn, each point refers to the addition of an increasing fixed amount of noise (noise is centered on zero and its standard deviation is, for each point, increased by 2). Finally, in our framework, each point corresponds to the normalization of a growing number of features (in order of increasing importance).

Results show that the suppression approach (slope: 0.12) seems the most advantageous in terms of compromise between utility and privacy. However it is very quickly limited by the number of selected features and therefore in privacy and utility metrics; for instance the best obtained performance are respectively 0.66 and 0.93. The perturbation approach (slope: 0.34) is very effective in loss of identification however at the cost of a very important loss of utility too, with for best performance in privacy and utility metrics respectively 0.51 and 0.84. Our approach is between the two (slope: 0.21) and provides the best utility and privacy trade-off (respectively 0.87 and 0.33).

Our approach based on normalization gives a better control on the weight of each feature in the protection, unlike the suppression approach which limits their impact to consideration or not.

Lastly, we also considered an adversary that trains a classifier only with features leading to the re-identification (Table III.3), in this case the accuracy in terms of re-identification is less efficient than with our framework (0.17).

III.1.5.3 Cost Improvement

We now compare the cost of running our framework with a traditional centralized solution based on an application server processing all the data. As described Section III.1.2.1, in a such solution, all the data collected by the IoT devices are sent to the application server

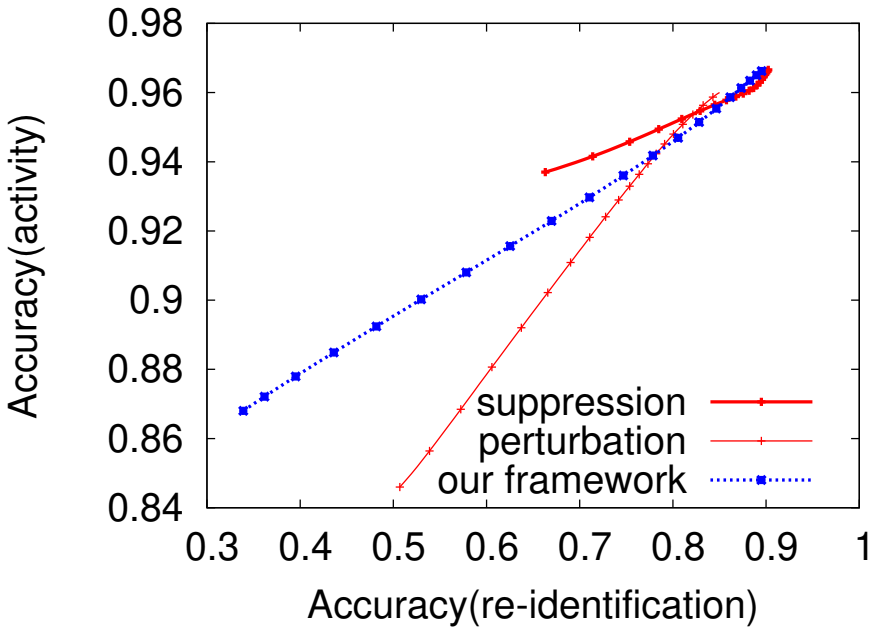


FIGURE III.11 – Our framework provides a better utility and privacy trade-off than baseline approaches.

which performs all the operations including signal preprocessing, segmentation, feature extraction, and classification as described Figure III.7. In comparison, our framework leverages the users smartphone to perform the signal preprocessing, the segmentation and the feature extraction, leaving to the application server only the classification task, it significantly reduces the computational cost of this server.

We measured the time spent by the application server for both the traditional approach and our approach. For our dataset (*i.e.*, 30 users and 15 minutes of data per user), the application server spent almost 52 seconds to perform all processing against almost 10 seconds with our framework (*i.e.*, classification only). That represents a time reduction of 81%. With a large number of users, this time reduction drastically saves the resources needed for operating an application server. Considering an application server running on a cloud infrastructure such as Amazon EC2 services [17], this reduction in terms of computational cost highlights the economic advantage of our framework.

We also measured the time spent by the application server when a homomorphic encryption scheme is used. Results show that using such a scheme to protect data generates important computational overhead. The added time introduced by the computation of the preprocessing (versus classification only) is negligible compared with this overhead.

Lastly, we evaluated the cost of running our framework on the user machine. A user only affords the cost of its activity by taking into account the preprocessing, the segmentation and the feature extraction on its smartphone. On a commodity computer, these operations applied on all the data of one user spent in average 2.5 seconds. This cost (paid every 15 minutes) remains low. In addition, these processing can be scheduled during the night when the user is inactive.

III.1.5.4 Impact of ML Scheme

To assess the generalization of results obtained by our approach with RF to other classifiers, we evaluate the utility-privacy trade-off with the classifiers considered in Section 2.5

(see Table 1). Table III.5 reports the privacy improvement and the utility loss of our approach according to the considered classifier.

Results depicted a similar behavior with an important decrease of the re-identification while maintaining the high level of activity recognition. There is an exception for Gaussian Naive Bayes classifier where the re-identification accuracy was already low before normalization. We also observe an increase of activity recognition after normalization for Support Vector Machine classifier probably due to the selection of features that significantly reduced the complexity of the data dimensionality that was too high on raw data for this basic classifier (*i.e.*, we go from 340 features to 19 features after application of our approach).

Algorithm	**Utility loss**	**Privacy improvement**
Decision Tree	-0.13 (0.81)	+0.54 (0.19)
K-nearest Neighbors	-0.07 (0.71)	+0.274 (0.09)
Support Vector Machine	+0.10 (0.68)	+0.16 (0.07)
Gaussian Naive Bayes	-0.16 (0.64)	+0.077 (0.06)
Random Forest	-0.09 (0.87)	+0.49 (0.33)
Quadratic Discriminant Analysis	-0.32 (0.56)	+0.555 (0.08)

TABLE III.5 – The tendency of our approach to drastically improve the privacy while maintaining the utility is generalized to other classifiers.

III.1.5.5 Impact of the Dataset

We now evaluate the capacity of our approach to be used on another dataset by learning a new model on this new dataset. In the previous experiments, the data acquisition protocol was done with volunteers wearing a smartphone on their waist. In this section, we use another dataset (*i.e.*, the "MotionSense Dataset" [199]) following a data acquisition done with a smartphone in the pocket of the volunteers. Consequently, the collected signals include more noise than in the previous dataset. Otherwise, this dataset is similar to the first one. It contains time-series data generated by 3-axial raw data from accelerometer and gyroscope sensors read at a constant frequency of 50 Hz. A group of 24 participants performed 6 different activities: dowstairs, upstairs, walking, jogging, sitting and standing. All these activities were made in 15 trials per user, with 9 long trials (*i.e.*, around 2-3 minutes duration) and 6 short trials (*i.e.*, around 30 seconds-1 minute duration).

Noise Sensibility To quantify the noise present in a collected signal, we measured the Signal-to-Noise Ratio (SNR for short). SNR is expressed in decibels as the ratio of the mean signal (μ) to the standard deviation (σ) of the signal over a given neighborhood using the logarithmic scale [49] :

$$SNR = 10\ log(\frac{\mu}{\sigma}).$$

This ratio compares the level of a desired signal to the level of background noise considering that the noise in the signal is stationary in time. A SNR higher than 0 indicates more signal than noise. Table III.6 reports the SNR for each activity for both datasets. Results show an important difference between the SNR of both datasets meaning a stronger presence of noise for all activities in the second dataset. This difference is significantly more important for statics activities (e.g., SNR at 25.4 versus 0.4 for standing activity). Indeed, the power of the desired signal for static activities is smaller, leaving more impact on noise.

Activity	Dataset 1 (smartphone on waist)	Dataset 2 (smartphone in the pocket)
Walking downstairs	5.05	3.8
Walking upstairs	6.6	0.1
Sitting	23.4	8.3
Standing	25.4	0.4
Walking	6.4	-3.0

TABLE III.6 – Signal-to-Noise Ratio (SNR) of collected signals of all activities for both datasets in decibels (dB) : the second dataset contains stronger noise on all activities, especially for static ones.

We now evaluate the impact of the presence of noise in the collected signal on the accuracy of our framework. Table III.7 reports the accuracy of the activity recognition. Results show that a Random Forest model is still able to highly recognise dynamic activities (*i.e.*, ranges from 0.79 to 0.89 of accuracy for walking and jogging activities) even if the collected data contains important levels of noise. However, the impact of noise on the signal of static activities drastically reduces the accuracy (e.g., 0.30 of accuracy for standing activity).

Activity	Accuracy (activity)
Walking downstairs	0.84
Walking upstairs	0.89
Sitting	0.16
Standing	0.30
Walking	0.80
Jogging	0.79

TABLE III.7 – Even if the collected data contains important level of noise, our framework is still able to highly recognise dynamic activities, while the impact of noise drastically reduces the accuracy for the recognition of static activities.

Figure III.12, in turn, depicts the cumulative distribution of the accuracy for the user re-identification task on this second dataset. This distribution shows that users can be still re-identified from their data even if the signals are perturbed by noise but with a smaller success rate than with the previous dataset (0.90 versus 0.48 of accuracy on average for the previous and the new dataset, respectively).

The impact of this noise can be mitigated by refining the preprocessing with an additional filter. For instance, we experimented adding a Savitzky–Golay filter that smooths the collected signal and consequently increases the SNR of each activity. This filter is based on the principle of fitting of an Nth degree polynomial to a set of input samples in a finite-length interval around the output sample time. It performs a low-pass filter using a local polynomial regression on the data sequence. By cutting-off the high frequency, we aim at keeping the global structure of the signals for example the cadence of the activities which are essential for their classification. Figure III.8 reports the accuracy of the activity recognition after applying a Savitzky–Golay filter, the performances has been increased with an average accuracy of 98%.

Feature selection To demonstrate the validity of our approach, it is important to ensure that the most important features used in the two classification tasks are mostly independent. Tables III.9 and III.10 lists for this new dataset the features selected for respectively

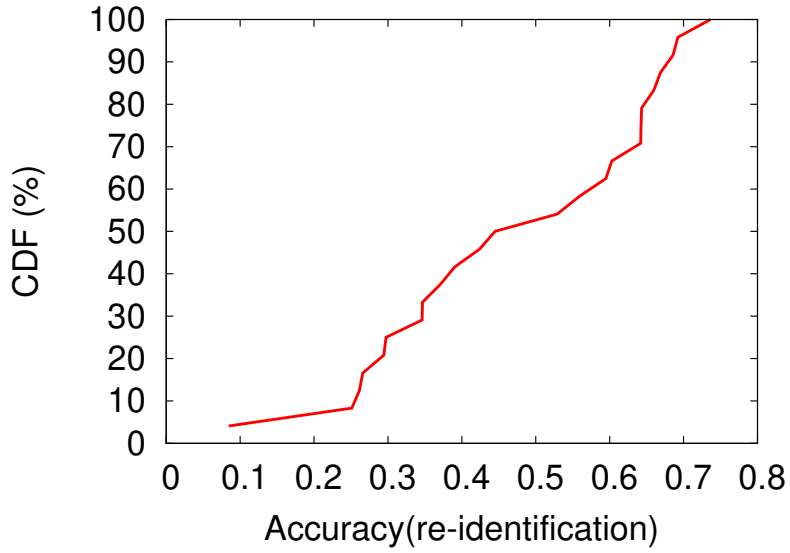


FIGURE III.12 – Cumulative distribution of the accuracy for the user re-identification task: users can be still re-identified from their data even if the signals are perturbed by noise but with a smaller success rate than with a dataset containing less noise.

Activity	**Accuracy (activity)**
Walking downstairs	0.95
Walking upstairs	0.92
Sitting	1.00
Standing	0.99
Walking	0.98
Jogging	0.97

TABLE III.8 – After applying a Savitzky–Golay filter, the classification accuracy for each activity has been increased

the re-identification and the activity classification, ranked following their importance in the classification task. Results show a majority of features in the frequency domain lead to the re-identification and a majority of temporal features lead to the activity recognition.

Framework analysis Our approach applied on MotionSense dataset provides on average an accuracy for activity recognition of 0.97 and an accuracy of 0.37 for re-identification. These results are similar to the ones obtained with the previous dataset (0.87 of activity recognition and 0.33 of re-identification). In summary, to get the best performances from our framework, the denoising has to be designed according to the SNR of raw data before applying the feature extraction. However, our framework still provides good activity recognition while reducing the users re-identification.

III.1.6 Conclusion

Our framework processes the signal and extracts relevant features locally on the user smartphone. In addition, accordingly to the observation that the frequency domain prevails in the user identification task, a normalization is performed on the frequency-based features to obfuscate the re-identification of users. Finally, only a set of features unlinked to the

Features	Importance
X_grav_mean	0.179
Y_grav_mean	0.153
Z_gyro_max	0.152
Z_grav_max	0.151
X_grav_max	0.138
Z_grav_max	0.113
Z_grav_var	0.112

TABLE III.9 – Features in the frequency domain also lead to the re-identification with the MotionSense dataset (frequency-based features are in grey)

Features	Importance
Y_gyro_mean	0.193
Y_body_std	0.184
Z_grav_std	0.167
Z_grav_mean	0.157
X_gyro_zcross	0.151
Magn_grav_min	0.148

TABLE III.10 – Temporal features also lead to the activity recognition with the MotionSense dataset (frequency-based features are in grey).

identity of its owner is uploaded to the application server which is then able to recognise the activity of the users with a high accuracy while reducing the risk of user re-identification. An extensive validation of our framework has been performed on 2 reference datasets yielding good results in terms of privacy-utility trade-off and suggesting that the approach could be generalized.

However, the different datasets were collected from a smartphone and it would therefore be necessary to evaluate our approach on data recorded via other mobile devices such as objects connected to the smartphone (e.g., smartwatch). Finally, despite these promising results, the method is limited to further decrease re-identification accuracy due to the limited number of features removable and also the potential decrease of activity accuracy if too many features are removed or normalized.

III.2 Motion sensor data anonymization by time-frequency filtering

The framework presented in the previous section allowed us to reduce the re-identification accuracy nevertheless the results were still far from the lowest achievable accuracy corresponding to a random guess $\frac{1}{n}$ with n the number of users. Furthermore, by minimizing information with a few number of features, the possibility to have a better trade-off than the one obtain in the previous section, is very reduced because the utility would be too much impacted. However, based on one of the main result concerning the feature discrimination between temporal and frequency for both activity and identity recognition, we could explore in this section a new method for anonymizing motion sensor data while preserving the remainder of the activity pattern.

As information patterns are contained in the temporal and in the frequency domain, our approach relies on time-frequency (TF) representation where sensitive information is removed to improve privacy. More precisely, since motion sensors respond to both frequency and intensity of movements and their outputs are non-stationary signals, a Fourier transform that provides frequency components of the whole signal is not sufficient to describe the signal properly. In this context, we propose to use a TF encoding of the signal to learn to recognize activity on one side and identity on the other. This learning step is based on a neural network which allows good learning performance and the automatic selection of descriptors of interest in the encoding space. Finally, we propose to filter sensitive identity information in this representation by canceling the highest value coefficients which correspond to the person's activity rate (see Figure III.13). The validity of our framework is extensively demonstrated on a public dataset and related to a state-of-the-art baseline.

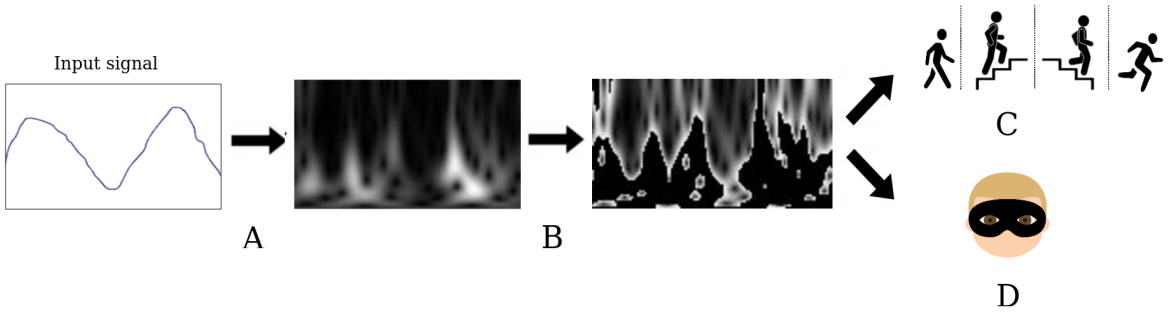


FIGURE III.13 – Overview of the proposed pipeline, divided in 4 steps: **A.** Signal transformation into a time-frequency (TF) image, **B.** Anonymization method based on image filtering, **C.** Activity recognition and **D.** User identification

III.2.1 Material and Method

III.2.1.1 Dataset

We used the public dataset Motion-Sense to assess the performance of our approach. It includes data sensed from 3-axis motion sensors at a constant frequency of 50 Hz collected with an iPhone 6s kept in the participant's front pocket [257]. Overall, a total of 24 participants have performed six activities during 15 trials in the same environment and conditions. The considered activities are going downstairs, going upstairs, walking, jogging, sitting and standing. To enable both classifications over time (*i.e.*, the activity and the identity), the raw

motion data are split in sliding windows, where each sliding window is a sample of a single activity. Knowing that on average the cadence range of walking is not less than 1.5 steps per second [37], the window length is chosen to be 2.5 seconds with an overlap of 50 %.

For this study, we focused on the four dynamic activities (going upstairs, going downstairs, walking and jogging): they are the most difficult to analyze and their complex frequency content is adapted to TF representation. Also, we only focused on the user acceleration signal, which is adapted for dynamic activities [254].

III.2.1.2 Time-frequency domain

The time-frequency domain tries to overcome the limitations of the classical Fourier transform, which only provides frequency content without any time information. Indeed, in the context of non-stationary signals, we need to know the frequency evolution of the signal components as a function of time. There are several approaches to project the signal in the TF domain. In this work we will focus on 3 different linear transforms: The Short-Time Fourier Transform (STFT) [107], the Stockwell transform [296] and the optimized Stockwell transform [217].

The Short-Time Fourier Transform (STFT) Given a signal $x(t)$, its STFT can be given as:

$$Sf_x(t,f) = \int_{-\infty}^{+\infty} x(\tau)w(\tau - t)e^{-2j\pi f\tau}d\tau, \tag{III.6}$$

where $w(t)$ is the analysis window which has a fixed width (mono-resolution analysis). The window $w(t)$ is chosen in this work as a Gaussian function with a standard deviation $\sigma = 0.05$.

The Stockwell transform The Stockwell transform (S-transform) can be considered as a hybrid between the STFT and the Continuous Wavelet Transform (CWT) [200]. It preserves a direct relation with the Fourier's kernel as the STFT, while performing a multiresolution analysis as the CWT. The S-transform of signal $x(t)$ can be expressed as follows:

$$S_x(t,f) = \int_{-\infty}^{+\infty} x(\tau)w(\tau - t,f)e^{-2\pi jf\tau}d\tau, \tag{III.7}$$

where $w(t,f)$ is the analysis window which is a Gaussian function of two variables: time and frequency. It can be given as:

$$w(t,f) = \frac{1}{\sigma(f)\sqrt{2\pi}}e^{\frac{-t^2}{2\sigma(f)^2}} \tag{III.8}$$

The standard deviation $\sigma(f)$ is inversely proportional to the frequency:

$$\sigma(f) = \frac{1}{|f|}, \tag{III.9}$$

to promote temporal resolution for high frequencies and frequency resolution for low frequencies.

Optimized S-transform To better adapt the analysis window to the nature of the signal being analyzed, many authors tried to optimize the S-transform representation by introducing new parameters on the Gaussian window [28, 282]. Among these proposals, a generalized Gaussian window controlled by a set of parameters was proposed in [217] as follows :

$$w^{p_i}(t,f) = \frac{|f|^r}{(mf^p+k)\sqrt{2\pi}} e^{\frac{-(\tau-t)^2 f^{2r}}{2(mf^p+k)^2}}. \tag{III.10}$$

The idea is to choose the set of parameters $p_i \in \{r, m, p, k\}$ that maximizes an energy concentration function. This energy concentration can be measured by several approaches. The concentration measurements (CM) used in this work are given as follows [294]:

$$CM(p_i) = \frac{1}{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|\overline{S_x^{p_i}(t,f)}\right| dtdf}, \tag{III.11}$$

with $\overline{S_x^{p_i}(t,f)}$ a normalization of $S_x^{p_i}(t,f)$ [282]:

$$\overline{S_x^{p_i}(t,f)} = \frac{S_x^{p_i}(t,f)}{\sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left|S_x^{p_i}(t,f)\right|^2 dtdf}}. \tag{III.12}$$

The parameters p_i which maximize $CM(p_i)$ are chosen to compute the optimized S-transform. In this study, the optimization is carried out on the whole signal and the optimal parameters were calculated for a sample of the signals in the dataset. This allows to observe the trend of the variation of these parameters particularly for the various activities where no significant variation is observed. In our case, the parameters are fixed as follows: $r = 0.7$, $m = 0$, $p = 0$ and $k = 0.4$.

III.2.1.3 Identity filtering

TF images generated from the different TF transformations have a size of 62 and 128 pixels in spectral (frequency voices) and temporal domains respectively. As depicted on Figure III.14, TF images for walking activity present different patterns from TF images for running activity, and can be discriminated in terms of texture: especially the number of vertical salient peaks. On the other hand, two different users can present differences in the contrast of their TF image as shown in Figure III.14, where peaks of user #8 are more contrasted than those of user #15. These observations stress the interest of filtering high coefficients to remove user information.

In agreement with these observations, identity filtering consists in setting different percentages x of the total TF image coefficients (sorted in descending order) to zero : x ranging from 10% to 90% with a step of 10%. This method allows us, in agreement with [157], to ensure that information relevant to re-identification is removed first.

III.2.1.4 CNN classifier

We propose two distinct convolutional neural networks as classifier to assess the performance of our framework: one that classifies signals into 4 classes (corresponding to the 4 activities) called CNN activity, and another that classifies the same signals into 24 classes (corresponding to the 24 subjects in the study) called CNN identity. These two models have

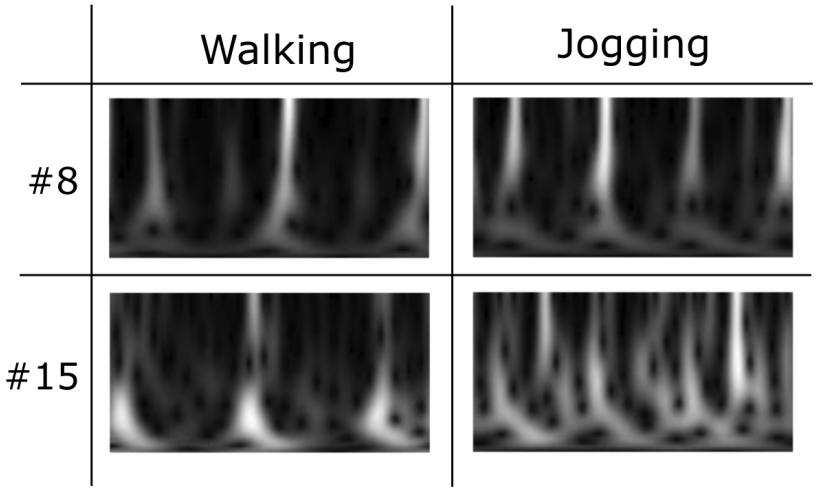


FIGURE III.14 – Representation of optimized S-transform for 2 different users (#8 and #15) for two different activities (walking and jogging).

the same architecture, but are trained separately.

CNN Architecture Multi-inputs image classification based on CNNs can be addressed using early fusion strategy, where all input images are combined at the beginning of the network. This fusion strategy present low computational complexity and is an easy implementation [214]. However, it has been shown in other contexts that the late fusion better accounts the complexity of each input and outperforms the early fusion [229, 293]. The late fusion strategy consists in processing each input image independently on distinct convolutional branches, and merging features at a higher level.

In our case, each signal window was defined by three different TF images, standing for the acceleration along the x, y and z axes (TF_x, TF_y, TF_z, respectively). The detailed architecture is depicted in Figure III.15.

CNN implementation For both CNN activity and CNN identity, signals in the dataset have been split according to trials: 90% of signals from trials 1,2,3,4,7,8,9 were used as training set, and the remaining 10% as validation set. Signals from trials 11,12,15,16 were used as testing set. For both CNNs, we used a categorical cross-entropy loss function that produced weights to equally penalize under or over-represented classes in the training set. The optimizer was set with Adam, the batch size was set to 128, and the number of epochs was set to 150 but was regulated by early stopping. The total number of weights to train was 23,044 for CNN activity and 46,104 for CNN identity.

III.2.2 Evaluation

III.2.2.1 Classification metric

To assess the classification performances of the two CNNs (activity and identity), we computed for each class the accuracy metric *acc*, defined as:

$$acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{III.13}$$

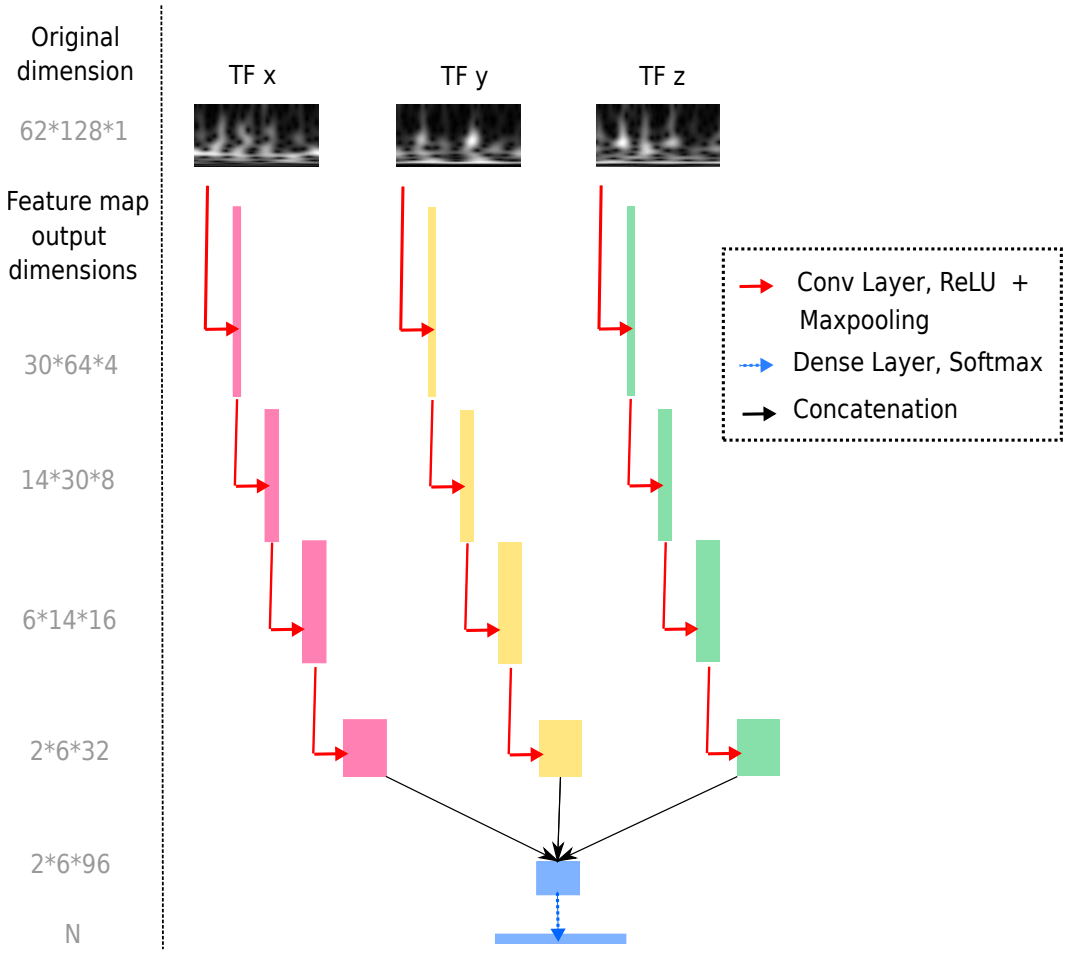


FIGURE III.15 – Overview of the proposed CNN architecture. The network takes three TF images (TF_x, TF_y, TF_z) as input. Each input image is processed independently on 3 separate branches. Pink, yellow and green feature maps result from 2D-convolutions and maxpooling. The output of the 3 branches are then concatenated, and passed through a hidden layer of N nodes, with $N = 4$ for CNN activity and $N = 24$ for CNN identity. The kernel size of the convolutional layers were defined as $3x3$.

where TP stands for true positives, TN for true negatives, FN for false negatives and FP for false positives.

We called the result respectively Activity *acc* (*i.e.*, data utility) and Identity *acc* (*i.e.*, privacy) when it is applied to the activity recognition and to the user identity. The given accuracies systematically corresponds to the results averaged over ten experiments.

III.2.2.2 State-of-the-art baseline

To compare the performance obtained from the different TF representations, we proposed a baseline based on the Fourier transform of the acceleration signal. In this baseline, just like our filtering approach, we filtered different percentages x of the transform coefficients in descending order (x ranging from 10% to 90% with a step of 10%). Once Fourier transform was filtered, as did [157], the signal was classified into activity and identity classes based on frequency domain features using a RF classifier.

III.2.2.3 Optimal representation metric

We defined the best TF representation as the one that maximizes the Area Under the utility-privacy Curve (AUC) using the trapezoidal rule. The AUC is an effective and combined measure of utility and privacy that describes the inherent validity of the anonymization approach. The AUC used here was bounded in x between the minimum and maximum performance in identity. As these bounds changed between the different TF representations, we normalized the AUC by the size of the rectangle relating to the bounds.

III.2.2.4 Optimal filter

The optimal filter was defined as the filter minimizing the Euclidean distance between a point from the utility-privacy curve and the upper left corner of normalized area – that corresponds to the intersection of the upper edge of the Figure III.16 (maximal performance of 100% in activity) and the minimum bound in performance for the identity. This optimal filter guarantees a good activity recognition while limiting identity identification.

III.2.3 Results and discussion

III.2.3.1 Interest of the time-frequency representation

Figure III.16 shows the filtering effect on activity and identity recognition for the three different TF representations and the baseline. It appears that working only in the frequency domain (cross markers in Figure III.16) leads quickly to a significant loss of activity recognition. This loss in data utility is explained by the fact that Fourier transform is not able to correctly analyze the non-stationary nature of signals, where frequency content changes over time especially in the activity pattern. On the other hand, TF representations seem to be able to deal with the non-stationarity of the signals and therefore allow a better trade-off between activity recognition and user identification. The more coefficients are filtered from the TF images, the worse identification performance is. Conversely, filtering has much less impact on activity recognition performance: whether no filtering is applied or that 70% of the image is filtered, activity accuracy seems to be stable between 80% and 90%. This trend is observed whatever the TF representation (STFT, S-transform or optimized S-transform respectively round, square and triangle markers in Figure III.16). These results demonstrate that high coefficients in the TF images carry specifically the person's activity rate and hence the identity information. Activity, on the other hand, seems less specific to a range of coefficients and corresponds more to the general texture observed in the TF images.

III.2.3.2 Optimal representation

Table III.11 summarizes the normalized AUC for each curve in the Figure III.16. Among the three TF representations, STFT has the lower AUC ($AUC = 0.83$), hence offering a worse trade-off between utility and privacy. This observation can be linked to the fact that this transform is mono-resolution and therefore provides a lower time-frequency resolution of the associated signal. On the other hand, the S-transform is multi-resolution which allows a better encoding of the analyzed signal and hence higher activity recognition ($AUC = 0.84$). Activity recognition is further improved when the S-transform of the acceleration signal is optimized according to an energy concentration criterion ($AUC = 0.85$). Differences between S-transform and optimized S-transform could possibly be even more marked if the S-transform optimization would have been done on individual sliding windows rather

TABLE III.11 – Normalized Area Under the utility-privacy Curves (AUC) for each representation

Fourier	**STFT**	**S-transform**	**Opti. S-transform**
0.69	0.83	0.84	0.85

than the whole signal. Unsurprisingly, the better is the energy concentration in the time-frequency plane – which means a better tonal resolution of the TF image – the faster the neural network converges.

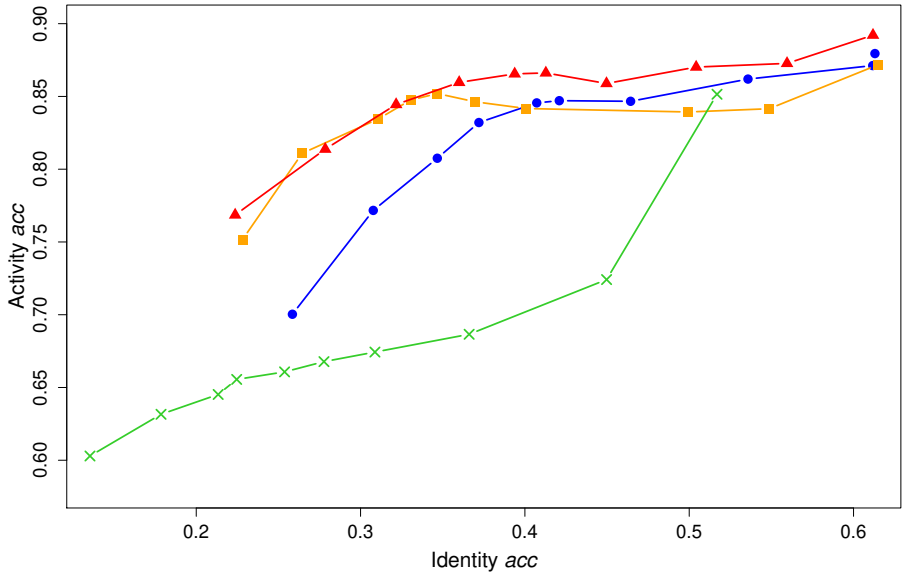


FIGURE III.16 – Activity accuracy according to identity accuracy for different representations: the Fourier transform (cross markers in green), the STFT (round markers in blue), the S-transform (square markers in orange) and the optimized S-transform (triangle markers in red). Each point corresponds to an average classification result over 10 experiments. The upper left corner represents the ideal trade-off between utility and privacy. For each curve, the high performance points in activity and in identity correspond to cases without filtering while the others (as one tends to the left of the graph) correspond to filtering cases with a step of 10%.

III.2.3.3 Optimal filter

Table III.12 shows that the optimal filter that guarantees a good utility-privacy trade-off is 60% for the Fourier, STFT and S-transform representations and 70% for the optimized S-transform representation. These observations suggest that given the better tonal resolution of the optimized S-transform representation, it is possible to filter more coefficients of the TF image without losing too significantly in activity performance (as observed for the other TF representations in the Figure III.16.

TABLE III.12 – Optimal filter in % for each representation, and the associated performances in % (activity *acc*/identity *acc*)

Fourier	**STFT**	**S-transform**	**Opti. S-transform**
60 (66/22)	60 (83/37)	60 (85/33)	70 (85/32)

III.2.4 Conclusion

In this work, we presented a new proof of concept method for preserving individual privacy in motion sensor data. This method uses time-frequency representation of acceleration signals and filters the resulting TF images by setting the highest coefficients to zero before the machine-learning step. The evaluations demonstrated that our method successfully anonymized identity, and preserved a high activity recognition ratio by better encoding of the non-stationary aspect of the signals than the Fourier transform. More specifically, we determined that the optimized S-transform gives the best utility-privacy trade-off by filtering its TF coefficients at 70%. The proposed filtering privacy-preserving mechanism was intentionally simple, but shows promising results. More advanced filtering methods [100] could be considered to improve performance, which could be the subject of future research. Moreover, other time-frequency transforms can be applied and compared with the results obtained in this contribution.

The contributions of this chapter focused on anonymization and the development of manual methods to prevent data from re-identification. These methods were based on the fact that information extracted from the temporal and frequency domains of the data are used either for activity recognition or re-identification. Thus, each framework focused on normalising and removing information related to re-identification in the form of features extracted from signals in the first contribution and TF images in the second contribution. However, the methods presented are specific to anonymization and then could not be applied on any other sensitive attribute.

Chapter IV

Sanitizing and FL scheme against inference attacks

This chapter aims at focusing on other sensitive inferences information that the user not necessarily consented to disclose and that can also be used as indirect re-identifiers by linking different information about the user on different datasets.

To do so, we firstly propose a solution based on deep neural networks that will automatically identify the features related to a specific sensitive information in order to obfuscate it in the motion sensor data. Secondly, we explore the distributed learning architecture and specifically we evaluate the privacy aspect of a FL scheme designed for heterogeneous data with private personalized layers.

IV.1 Dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks

In this work we propose a solution sanitizing the motion sensor data in such a way that it hides sensitive attributes while preserving the activity information contained in the data.

To achieve this objective, we designed DySAN, inspired from the framework of GAN [238] to sanitize the sensor data. By learning in a competitive manner several networks, DySAN is able to build models for sanitizing motion data to prevent inferences on a specified sensitive attribute while maintaining a high level of activity recognition. One of the objectives of DySAN is also to limit the distortion between the raw and sanitized data, thus also maintaining a high level of utility with respect to other analysis tasks related to activity monitoring (*e.g.*, steps counting). DySAN has thus to learn sanitizing models in order to find the best trade-off to deal with these conflicting optimizing goals.

Furthermore, our approach aims at addressing the heterogeneous aspect of data collected by motion sensors. Indeed, these sensor data are user dependent and inherently reflect the way each user moves, to the characteristics of the sensors used for data collection and to the evolution of activity during the day. Thus, one unique sanitizing model cannot cope with the heterogeneity of data and provide the best utility/privacy trade-off for all users over time. To solve this issue, DySAN builds a set of diverse sanitizing models by exploring different combinations of hyperparameters leading to different balance in privacy protection with respect to sensitive inference, utility preservation in terms of the loss induced for activity recognition, and the data distortion. By doing so, DySAN is able to assess the trained sanitized models and to dynamically select the model providing the best trade-off over time according to the incoming sensor data.

The evaluation of DySAN on real datasets, in which the *gender* is considered as the sensitive information to hide due to the possible risk of discrimination, demonstrates that DySAN can drastically limit the gender inference with a drop of 41% while only inducing a drop of 3% on the accuracy of activity recognition. In addition to preserving activity recognition, DySAN, by limiting data distortion, also preserves the sensor data utility for other analytical tasks such as estimating the number of steps. Moreover, we show that the dynamic

model selection of DySAN successfully provides an adaptation of the sanitization according to the incoming user data. This dynamic model selection is especially useful to generalize the sanitization capacity learnt from the dataset used to build the sanitizer models to another dataset with new users with potentially different behaviours. Our dynamic sanitizing method overcomes several shortcomings of the state-of-the-art approaches, namely the use of the same sanitizing model for all users over time, which may lead to a poor privacy-utility trade-off for atypical users. Lastly, we evaluate the cost of operating DySAN on a smartphone and show that the introduced overhead is compatible with real-time processing and that the energy consumption remains reasonable.

IV.1.1 Problem definition and system model

IV.1.1.1 Overview and system model

The system configuration is based on similar hypothesis to those presented in the previous chapter. We consider a mobile application installed on the user's smartphone aiming to monitor its physical activity. The smartphone of the user is assumed to be under the control of the user and thus trusted while the service provider responsible for activity monitoring is not. In the "classical" (*i.e.*, non-private) version of activity monitoring, the data is acquired by sensors (*e.g.*, accelerometer, gyroscope), retrieved by a dedicated mobile application and then send to a server hosted in the cloud (*i.e.*, no classification is made locally). This server leverages ML models to identify the activity of the user or to estimate other physical activity features (*e.g.*, number of steps). The server is considered to follow the honest-but-curious adversary model in the sense that it may also try to infer additional sensitive information from the sensor data.

For the rest of the section, we consider the *gender* as being the sensitive attribute to protect as it could lead to risk of discrimination. In addition, the gender could be inferred from the list of performed activities and their associated frequencies in case of unbalanced data distribution between men and women (which is not the case in the datasets considered in this section). However, our approach is much more generic and could be applied to protect other sensitive attributes (*e.g.*, handicap or race).

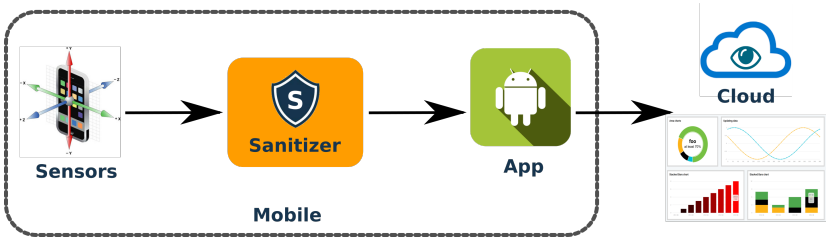


FIGURE IV.1 – DySAN locally sanitizes the motion sensor data on the smartphone to prevent the cloud-based service from inferring an unwanted sensitive attribute while allowing this service to detect the activity performed by the users as well as compute statistics related to their physical activity.

An overview of DySAN is shown in Figure IV.1. To avoid an unwanted exploitation of the motion sensor data, these data are sanitized by DySAN before being transmitted to the mobile application. This sanitizing process removes the correlations with the sensitive attribute in the sensor data while preserving the information necessary to detect the activity performed by a user. In addition, DySAN also aims at limiting the distortion between the raw and sanitized data to preserve the utility for other analytical tasks. Finally, the resulting

sanitized data are sent to an analytics application hosted in the cloud, exploiting ML models to classify the users activity and compute statistics related to their physical activity.

Ideally to limit the privacy risks, the sanitizing step should be done as early as possible to prevent other applications to have access to the raw sensor information. For instance, we consider that DYSAN could be deployed in the trusted environment of the smartphone to prevent the mobile application to have a direct access to the sensor data but only from the output of DYSAN (thus ensuring that the mobile application uses only sanitized data). Afterwards, the mobile application sends the sanitized data to a server hosted in the cloud similarly to the non-private scenario.

IV.1.1.2 Problem statement

More formally, we consider raw motion sensor data (denoted by A) captured through accelerometer and gyroscope that sample 3-axial signals with a frequency of 50 Hz. To enable activity recognition over time, the raw sensor data are split in sliding windows, in which each sliding window is considered to be a sample of a single activity (*i.e.*, by assumption the user cannot perform two different activities during a single sliding window). The choice of the window size is not trivial, especially for an activity recognition task and has to be well calibrated. Indeed, a small window size could split an activity signal while a large window size could contain multiple activity signals. Knowing that on average the walking pace is not less than 1.5 steps per second [37], the window length T is chosen to be 2.5 seconds with an overlap of 50 % to match that of a walking cycle of two steps.

We assume a population of N users contained in a dataset X storing all user's data. This dataset includes the raw sensor data as well as the label associated with the activity performed by the user (denoted by a multi-valued attribute Y), the binary sensitive attribute (denoted by S) and a timestamp. Thus, the dataset $X = \{A, Y, S\}$ in which $A = (A_1, \ldots, A_T)$.

The objective of DYSAN is to protect the user motion sensor data against sensitive attribute inferences while maintaining data utility. More formally, we aim at learning a set of sanitizers $S_{an_{\alpha,\lambda,\beta}}$ for various values of the hyperparameters α, λ and β. Each sanitizer will transform the original data X into $\bar{X} = San_{\alpha,\lambda,\beta}(X) = \{\bar{A}, Y, S\}$; $\bar{A} = (\bar{A}_1, \ldots, \bar{A}_T)$. This set of sanitizers is learned so that it is difficult to build a discriminator D_{isc} trained to predict S from the sanitized data and activities $\{\bar{A}Y\}$ while an activity predictor P_{red} trained on the same sensor data ($\bar{A}$) is able to maintain an accuracy close to the original one (more details in the following section). To preserve further the utility of $\bar{X}$, the sanitizing process is constrained to minimize the distortion between the original and sanitized data.

Furthermore, DYSAN aims to dynamically adapt over time the hyperparameters of the model according to the incoming data of each user. Indeed, while a particular model could provide the best utility/privacy trade-off on average for all users with respect to the training dataset, the model leading to the best trade-off can change when testing on new user (*e.g.*, when the new user data does not fit the data distribution of the training dataset). More formally, the objective is to find for each window of data the sanitizer $\widehat{San_{\alpha,\lambda,\beta}}$ providing the best utility/privacy trade-off for the current incoming data. This trade-off is defined by a metric combining the accuracy of the activity recognition and the inference of the sensitive attribute.

IV.1.2 Dynamic Sanitizer

Before exploiting DYSAN, multiple sanitizers corresponding to various utility and privacy trade-offs are built during the offline training phase. These models are then deployed on the smartphone. During the online phase, DYSAN dynamically selects on the smartphone

the best sanitizer for the incoming sensor data. Both the training and the online phases are summarized in Figure IV.2 and explained in the following subsections.

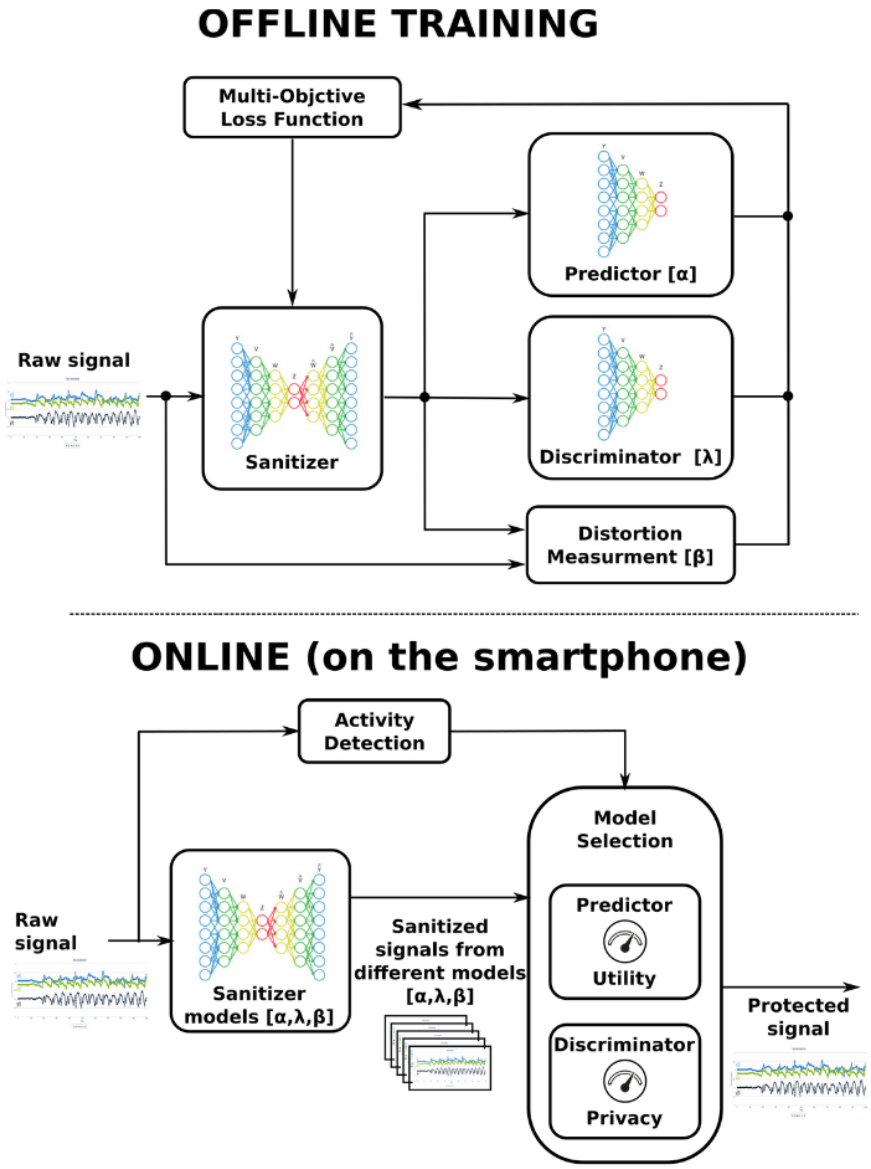


FIGURE IV.2 – DYSAN is composed of two phases: an offline training phase (left) and an online phase (right). The training phase is performed only once and aims to build different sanitizer models that are distinguished by their hyperparameters. Once these sanitizer models deployed on the smartphone, the online phase aims to dynamically choose among these models the most adapted one for each batch of incoming data.

IV.1.2.1 Building multiple sanitizers

The offline training phase is performed only once and aims at learning multiple sanitizers. This training is performed with a reference dataset used in activity recognition, the MotionSense dataset that we describe in Section IV.1.3.1. As shown in Figure IV.2, DYSAN is composed of multiple building blocks that we detail hereafter: 1) a sanitizer, 2) a discriminator, 3) a predictor, 4) a distortion measurement and 5) a multi-objective loss function.

— **Discriminator:** The discriminator Disc guides the sanitizer through the process of removing information related to the sensitive attribute $S \in \{0, 1\}$. In practice, we use a CNN, which is well-suited to capture time-invariant features in time series [147]. This CNN is described by the architecture in the following Table IV.1.

TABLE IV.1 – Discriminator architecture

Type	Parameter
Input	125,6
Conv1D	64, kernel_size=6, stride=1, activation=ReLU
AvgPool1D	kernel_size=2, stride=2
BatchNorm1D	100, eps=1e-05, momentum=0.1
Dropout	p=0.5
Dense	64, activation=ReLU
Dense	2, activation=softmax

The training of the discriminator is based on a loss function measuring the Balanced Error Rate BER [95] between the output of the discriminator and the ground truth sensitive attribute, which is defined as:

$$BER(Disc(\bar{A}, Y), s) = \frac{1}{2}(\sum_{s=0}^{1} P(Disc(\bar{A}, Y) \neq s | S = s)). \tag{IV.1}$$

The value of BER ranges between 0 and 0.5, in which a value close to 0 corresponds to a perfect accuracy for the prediction of the sensitive attribute while 0.5 means the discriminator is unable to retrieve any information about the sensitive attribute from the sanitized data. Hereafter, we will refer to this loss by $Loss_{Sensitive}$.

— **Predictor:** The predictor Pred aims at helping the sanitizer in preserving as much information as possible with respect to the activity recognition task. We also use a CNN for the predictor that has been optimized for predicting the user activity from the sanitized data. The architecture of this CNN is presented in the following Table IV.2.

TABLE IV.2 – Predictor architecture

Type	Parameter
Input	125,6
Conv1D	64, kernel_size=6, stride=1, activation=ReLU
AvgPool1D	kernel_size=2, stride=2
BatchNorm1D	100, eps=1e-05, momentum=0.1
Conv1D	100, kernel_size=5, stride=1, activation=ReLU
AvgPool1D	kernel_size=2, stride=2
Conv1D	160, kernel_size=5, stride=1, activation=ReLU
AvgPool1D	kernel_size=2, stride=2
Conv1D	160, kernel_size=5, stride=1, activation=ReLU
AvgPool1D	kernel_size=2, stride=2
Dropout	p=0.5
Dense	64, activation=ReLU
Dense	4, activation=softmax

Thus, the predictor is trained to maximize the accuracy in inferring activities from the output of the sanitizer. We also use the balanced error rate as the loss function that should minimize the error between the output of the predictor and the ground truth of the activity: $BER(Pred(\bar{A}), y)$. For the rest of the section, we will refer to the predictor loss as $Loss_{Activities}$.

— **Distortion measurement:** The last constraint on the sanitizer is the minimization of data distortion between the raw and sanitized data. Specifically, this distortion should be limited to keep as much information as possible in the sensor data for subsequent analytical tasks. The data distortion is measured through the L_1 loss function denoted *l1*, applied independently on each attribute. For two vectors A_i and $\bar{A}_i$, corresponding respectively to the raw and sanitized sensor data, the loss function is defined as follows:

$$l1(A_i, \bar{A}_i) = \frac{1}{N_A} \sum_{j=1}^{N_A} |a_{ij} - \bar{a}_{ij}|, \tag{IV.2}$$

in which N_A is the number of possible values for a particular attribute (*e.g.*, the number of axes of the accelerometer or the gyroscope), $a_{ij} \in A_i$ and i denotes a single observation in the window of length T.

— **Sanitizer:** The sanitizer San modifies the raw data taken as input to remove information correlated with the sensitive attribute while maintaining useful information for activity detection. Since the raw and sanitized data belong to the same space, we have implemented the sanitizer as an auto-encoder. In a nutshell, an auto-encoder is a neural network performing a dimension reduction of the signal to compress information before trying to reconstruct the input. The sanitizer takes into account the feedback of the discriminator, predictor and distortion measurement to output the sanitized version of the input raw data. More precisely, these different feedbacks are integrated into a multi-objective loss function that should be minimized. The architecture of the auto-encoder is given in the following Table IV.3.

— **Multi-objective loss function** The multi-objective loss function J^{San} drives the transformation performed by the auto-encoder to generate the sanitized data $\bar{X}$. This loss function takes into account three components, the capacity to detect the activity of the user (*i.e.*, the output of the predictor), the capacity to detect the sensitive attribute (*i.e.*, the output of the discriminator), and the level of distortion introduced in the sanitized data compared to the original one. More formally, the multi-objective is defined as follows:

$$\begin{aligned} J^{San}(X,\ San,\ Disc,\ Pred) = \{&\alpha * d_s(S, Disc(San(X))), \\ &\lambda * d_p(Y, Pred(San(X))), \\ &\beta * d_r(X, San(X))\}, \end{aligned} \tag{IV.3}$$

in which $d_s(x) = \frac{1}{2} - Loss_{Sensitive}$, $d_p = Loss_{Activities}$ and $d_r = \{l1(a_{:,j}, \bar{a}_{:,j}), \ldots\}$ with $a_{:,j}$ representing a dimension of all timesteps of a single sliding window. The term $\frac{1}{2}$ in $d_s(x)$ comes from the objective of maximizing the error of the discriminator, since the sanitizer aims at modifying the data so that the discriminator is no more able to infer sensitive information.

A gradient descent is applied on J^{San} to minimize the global loss function following a similar approach as in [10]. Note that each loss term is weighted with a hyperparameter. More precisely, d_s, d_p and d_r are weighted respectively with α, λ and β. The parameter α represents the relative importance given to the privacy while λ controls the utility (*i.e.*, the quality of activity detection). As we impose the constraint that $\alpha + \lambda + \beta = 1$, we only adjust α and λ hyperparameters, leaving $\beta = 1 - (\alpha + \lambda)$.

TABLE IV.3 – Sanitizer architecture

Type	Parameter
Input	125,6
Conv1D	64, kernel_size=6, stride=1
Conv1D	128, kernel_size=5, stride=1
Dense	128
Dense	64, activation=LeakyReLU(0.01)
Dense	64
Dense	128
Deconv1D	128, kernel_size=5, stride=1
Deconv1D	64, kernel_size=5, stride=1

IV.1.2.2 Training Phase

During the training phase, we build a sanitizer for each set of possible values for the hyperparameters α and λ to explore the domain of the multi-objective loss function. This exploration will allow DYSAN to select the best model for each user during the online phase. The training procedure is summarized in Algorithm 2.

In order to optimize the utility and privacy trade-off for a specific set of α and λ (line 1, Algorithm 2), the three neural networks are trained in an adversarial manner. This adversarial training can be seen as a game between the sanitizer on one side and the predictor and the discriminator on the other side. These neural networks compete against each other with opposing objectives until an equilibrium is reached. More precisely, the sanitizer is trained to fool the discriminator and maintain a high activity detection quantified with the predictor while limiting the data distortion. We follow the standard training procedure of GANs consisting in alternating in an iterative manner (at each batch of data) the training of each model with their respective loss function until convergence or until a maximum number of epoch (*i.e.*, we do not consider Competitive Gradient Descent [276]).

Specifically, after initialization (lines $1-6$) the training of the sanitizer starts with J^{San} while the discriminator and the predictor are frozen (lines $9-10$). Once the training of the sanitizer has converged, the predictor and the discriminator are trained independently with their respective loss function while the sanitizer is frozen (lines $11-18$). These two neural networks are trained until convergence (*i.e.*, until the loss no longer decreases) or if a maximum number of iterations, respectively K_{pred} and K_{disc}, is reached. This two-steps process is performed iteratively until an equilibrium is reached.

IV.1.2.3 Online Phase

Once deployed on the smartphone, DYSAN is composed of four components as depicted in Figure IV.2: the sanitizer, the discriminator, the predictor and an activity detection component. Specifically, DYSAN knows all the sanitizer, predictor and discriminator models built

Algorithme 2 : DySAN training algorithm

Input : X, λ, α, max_epoch, $batch\_size$, K_{pred}, K_{disc}
Output : San, $Disc$, $Pred$

train(M, **trParams): Train the model M using trParams.
freeze(M): Freeze the model M parameters and avoid modifications.
Initialisation
San, $Disc$, $Pred$, $X_d = \text{shuffle}(X)$, $X_p = \text{shuffle}(X)$
$\text{Iterations} = \frac{|D|}{\text{batch\_size}}$
Training Procedure
for $e = 1$ ***to*** max_epoch **do**
  for $i = 1$ ***to*** Iterations **do**
    Sample batch B of size batch_size from X
    train(San, B, J^{San}, α, λ, freeze($Pred$), freeze($Disc$))
  end
  for $k = 1$ ***to*** K_{pred} **do**
    Sample batch B of size batch_size from X_p
    train($Pred$, B, $Loss_{Activities}$, freeze(San))
  end
  for $k = 1$ ***to*** K_{disc} **do**
    Sample batch B of size batch_size from X_d
    train($Disc$, B, $Loss_{Sensitive}$, freeze(San))
  end
end

during the training phase. This set of models corresponds to the different possible utility and privacy trade-offs (*i.e.*, set of values explored for the α and λ hyperparameters). The selection of the model is performed by maximizing $S(P, U) = xU + yP$, in which x and y are positive weight coefficients with $x + y = 1$, U the evaluation of the activity done by the predictor, and P the accuracy in terms of privacy as $P = 1 - |0.5 - p|$, in which P is the evaluation of the gender done by the discriminator. Consequently, P is higher when the evaluation of the gender accuracy corresponds to a random guess (*i.e.*, an accuracy of 0.5). According to the expected utility and privacy trade-off, the coefficients x and y can be tuned (the impact of these parameters is assessed in Appendix IV.1.4.2).

To find the best sanitizer over time (according to coefficients x and y), DySAN evaluates the utility and the privacy of all models to select the best one. This evaluation requires to know the actual activity performed by the user and the sensitive attribute. While the sensitive attribute can be given by the user, the motion sensor data are not labeled with the activities as it is rather the objective of the activity recognition task to perform this inference.

We use the activity detection component (see Figure IV.2) to annotate some motion sensor data with their activities on the smartphone. More precisely, we ask the user to follow a specific calibration process at the installation of DySAN. During this process, the user is asked to perform a series of different activities for short periods to learn a specific supervised classifier to detect his activities. As the quantity of data available to train this classifier is limited, we rely on the use of RFs that are adapted to this context [157]. This RF classifier is then used to label the raw data in order to evaluate the utility and the privacy of all sanitizers. This evaluation is performed on a regular basis (*e.g.*, each period of p windows) and we compute the average accuracy over this basis. By following this process, DySAN is

able to identify over time the sanitizer providing the best utility and privacy trade-off defined as a measure combining the accuracy of the activity recognition and the inference of the sensitive attribute. DySAN with this specific online process is called DySAN (o) in the next evaluations.

IV.1.3 Experimental setting

IV.1.3.1 Datasets

We used two real datasets, which are both publicly available and heavily used in the literature: MotionSense and MobiAct. These datasets contain motion sensor data of subjects doing cyclo- stationary activities (*i.e.*, based on step pattern).

— **MotionSense** [257] contains data captured from an accelerometer (*i.e.*, acceleration and gravity) and gyroscope at a constant frequency of 50Hz collected with an iPhone 6s kept in the front pocket. Overall, a total of 24 participants have performed six activities (*i.e.*, going downstairs, going upstairs, walking, jogging, sitting and standing) during 15 trials in the same environment and conditions.

— **MobiAct** [317] records the data from 58 subjects during more than 2500 trials, all captured with a smartphone in a pocket. This dataset includes signals recorded from the accelerometer and gyroscope of a Samsung Galaxy S3 smartphone with subjects performing nine different types of activities of daily living. For our experiments, we only used the trials corresponding to the same activities as MotionSense.

Both datasets are balanced and contain an equal number of males and females. However, the walking activity is more represented compared to others (Section IV.1.4.4). Each activity is performed equally by all subjects in both datasets, making any correlation between the gender and the activity impossible. The datasets are split between training and testing, with 2/3 of trials used for training and validation and 1/3 for testing. These two datasets share similar characteristics, which allows to test the transferability of the models from one dataset to the other. More precisely, the models learned on one dataset can be used to sanitize data from the other dataset. This evaluation corresponds to a more realistic use case and to the best of our knowledge was never considered in previous work related to the sanitization of sensor data.

IV.1.3.2 Baselines

To assess the performance of DySAN, we considered a set of baselines that we detail hereafter. One of these baselines is based on a RF classifier [157] while the others are based on GANs [197, 199, 253]. Regarding GAN approaches, authors use an architecture of neural networks slightly different from ours. To provide a fair comparison, we propose to implement their functionalities in our architecture (number of layers, type of CNN, ...). This methodology allows us to assess the main characteristics adopted in the baselines without depending on their choice of architecture that can also have an impact on performance.

— **ORF:** It corresponds to the first the contribution in SectionIII.1. The raw data is preprocessed on the user's smartphone and only relevant features are transmitted to the application hosted in the cloud. The relevant features are first identified according to the target application (*e.g.*, activity recognition) and selected either in the temporal or the frequency domain. Originally proposed to avoid users re-identification, we adapt this approach to prevent the inference of the sensitive attribute, namely gender. More specifically, we first detect the features that are the most correlated with the gender

before normalizing the features in the frequency domain and removing the features in the temporal domain that are not used for the activity classification.

— **GEN:** Similarly to DySAN, GEN (Guardian Estimator Neutralizer) [199] also relies on an adversarial approach to optimize the utility and privacy trade-off. However, this solution does not follow the standard iterative training procedure of GANs as described in Section IV.1.2.2. More precisely, the first network, a classifier, is learned once on the raw data to identify both sensitive (*e.g.*, the gender) and non-sensitive information (*e.g.*, the activity). This combination is made using a Multi-task neural network. Then the second network, an auto-encoder, is also trained only once through a loss function that does not take into account the data distortion. Finally, the model used in the online phase is the same for all users and corresponds to the best set of hyperparameters identified during the training phase. While this solution relies on a neural network architecture slightly different from ours, we implement GEN by using our architecture. However, to evaluate the performance of GEN in a context of transfer learning, we also use their original neural networks (learned on MotionSense [1]) to assess its performance on MobiAct.

— **Olympus:** This approach [253] is similar to GEN with the exception that two different neural networks are used to learn the sensitive attributes and to learn non sensitive information. In addition, these classifiers are trained using sanitized data by following an iterative process similar to DySAN described in Section IV.1.2.2. However, the loss function does not account for data distortion and the model deployed is the same for all users (*i.e.*, only one sanitizing model trained and used in the online phase). While this approach is used for a different objective (*i.e.*, to avoid users re-identification), we adapt it for activity recognition by using our architecture.

— **MSDA:** This solution [197] can be viewed as an evolution of Olympus in which the loss function driving the training of the auto-encoder accounts for data distortion. However, the model used in the online phase is still the same for all users. While this approach was originally developed with a different purpose in mind (*i.e.*, to avoid re-identification), we adapt this solution for activity recognition by using our architecture. This baseline is the closest to DySAN but without the training of multiple sets of hyperparameters and the dynamic sanitizing model selection in the online phase.

As described in Section IV.1.2.3, in the online phase DySAN selects the sanitizing model which provides the best utility and privacy trade-off controlled by parameters x and y. The reported results in the evaluation correspond to an unbalanced trade-off to improve privacy (*i.e.*, $x = 0.1$ and $y = 0.9$). The impact of different parameters is assessed in Section IV.1.4.2.

IV.1.3.3 Evaluation metrics

We evaluated DySAN along both utility and privacy metrics, and a couple of system-level metrics.

— **Utility:** In our context of physical activity monitoring, the first considered utility metric is the accuracy of a classifier for activity recognition. More precisely, we use the confusion matrix derived by this classifier to measure the number of correct predictions made by the classifier over all predictions made. The value of the accuracy ranges from 0 to 1, in which 1 corresponds to perfect accuracy. In addition, analytics applications monitoring physical activity usually compute and present many estimators to users. To evaluate this aspect, we compute the number of steps detected from the sanitized

1. https://github.com/mmalekzadeh/motion-sense/tree/master/codes/gen_paper_codes

data and compare it with the number of steps in the raw data. To realize this, we first normalize the raw and sanitized data to compare them in the same range of values, and then compute a Peak Acceleration Threshold [3] from the raw data to estimate the number of peaks. More precisely, we used Adaptiv: An Adaptive Jerk Pace Buffer Step Detection Algorithm (`https://github.com/danielmurray/adaptiv`) for estimating the number of steps detected by the analytics application from the received data.

— **Privacy:** To assess the level of privacy of DySAN, we rely on the accuracy of inferring the sensitive attribute. In our case, we consider the gender of the users as the sensitive attribute (personal information available in public datasets). An accuracy of 0.5 corresponds to a random guess as our dataset is balanced.

— **System-level:** To assess the overhead of operating DySAN on a smartphone, we measure both the CPU time spent to sanitize the raw data on the smartphone and the energy consumption over time during a real-time processing of DySAN.

IV.1.3.4 Methodology

DySAN is trained only with the MotionSense dataset while the results reported for MobiAct evaluate the transfer learning (*i.e.*, using sanitizing models trained on MotionSense to sanitize data from MobiAct). In the training phase, we explore a range of values between 0.1 and 0.9 with a 0.1 step for both α and λ, which corresponds to 36 different sanitizing models. The sanitizer models of DySAN are trained for 300 epochs and the size of a data batch is set to 256 samples. In the online phase, we select a privacy and utility trade-off focusing primarily on privacy (*i.e.*, ensuring the protection of the gender at the cost of the accuracy). This trade-off is controlled by the parameters x (utility) and y (privacy) (Section IV.1.2.3) which are set respectively to 0.1 and 0.9.

The RF classifier applied during the online phase of DySAN uses a feature vector extracted from the raw signal. The choice of these descriptors was made on the basis of an earlier review on effective descriptors for gait recognition [290]. The data of each user is splitted into train, validation and test subset in order to have all the users in each subset. The data firstly splitted into 75% and 25% for train/validation and test subsets. Then splitted into 80% and 20% for train and validation. We use 4-fold cross-validation in which the testing set is randomly partitioned into 4 equal sized subsamples. Reported results correspond to average over 10 repetitions of each experiment. The computation of the different global models (each corresponding to a precise set of hyperparameters) has been parallelized on a hybrid GPU/CPU computing farm.

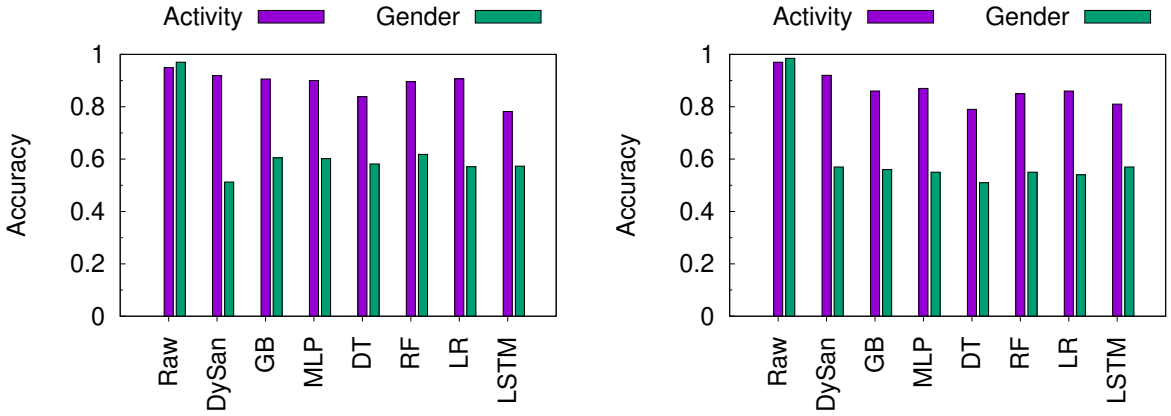
IV.1.4 Evaluation

In this section, we report the results obtained for the evaluation of DySAN by highlighting important features, namely the good utility and privacy trade-off (Section IV.1.4.1), the low distortion of the sanitized data (Section IV.1.4.3), the better performances compared to state-of-the-art approaches (Section IV.1.4.4), the advantage of dynamically select the best sanitizing model according to the incoming data (Section IV.1.4.5), and the limited cost of operating DySAN on a mobile (Section IV.1.4.6).

IV.1.4.1 Utility and privacy trade-off

In this section, we evaluate the capacity of an analytics application to infer the gender of the user and its activity from the sanitized data provided by DySAN and sent by the mobile application. We compare the performance of several classifiers that could be used by the

analytics application, namely a gradient boosting classifier (GB), a multi-layer perceptron (MLP), a Long Short-Term Memory (LSTM), a decision tree (DT), a RF, a logistic regression (LR) and also two CNNs with the same architectures than the predictor and the discriminator of DYSAN (referred as Raw on the figures).



(a) MotionSense (b) MobiAct

FIGURE IV.3 – The sanitized data provided by DYSAN drastically decreases the privacy risk compared to using the raw data while limiting the loss of activity detection, and this regardless of the classifier used.

Figure IV.3 reports the accuracy for both datasets for predicting the gender and the activity with the different classifiers as well when using the raw data. First, the results show that without any protection (*i.e.*, on raw data) the application is able to infer the gender with 98.5% accuracy. In addition, the activity is also inferred from the raw data with 97% of accuracy on average. Secondly, we can observe that DYSAN successfully decreases the privacy risk with respect to inferring the sensitive attribute while limiting the loss of activity detection. Indeed, with the sanitized data, an analytics application is only able to infer the gender up to 61% and 57% of accuracy, respectively for MotionSense and MobiAct. In terms of utility, depending on the classifier, the accuracy of the activity recognition varies between 78% and 92%, which represents only a small drop compared to using the raw data. Remark that the LSTM, a recurrent neural network architecture commonly used for temporal signal, does not provide best results as one could expect.

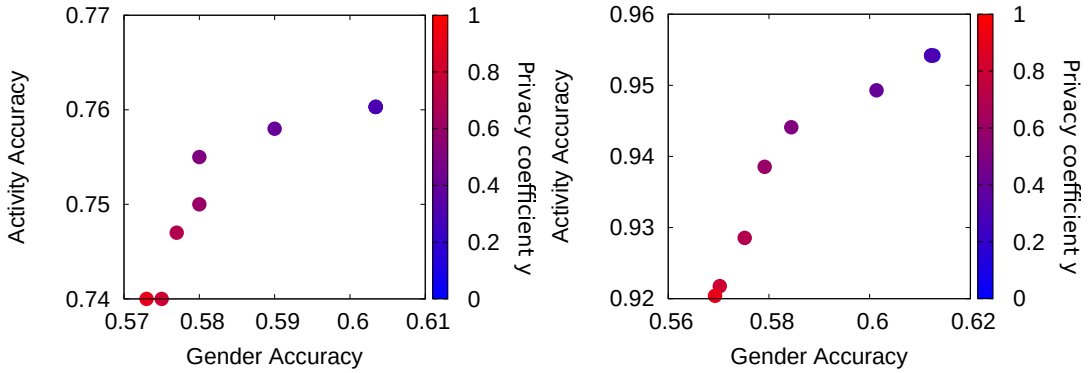
IV.1.4.2 Impact of weight coefficients on DYSAN

As described in Section IV.1.2.3, the best sanitizer model is selected according to the definition of the utility and privacy trade-off defined by weight coefficients x and y. Figure IV.4 depicts the evolution of the utility and privacy trade-off according to x and y for both datasets.

For both dataset, when y increase, the Privacy increase (the gender accuracy decrease) and the Utility decrease (the activity accuracy decrease).

IV.1.4.3 Distortion of the sanitized signal

The utility of the sanitized data is not just about the activity recognition but also with respect to more fine-grained information related to the activity. In this section, we demonstrate that DYSAN keeps relevant information in the signal enabling us to conduct further analysis. More precisely, we consider the computation of the number of steps from the signal for MotionSense dataset. Following the step detection method presented in IV.1.3.3,



(A) MotionSense (B) MobiAct

FIGURE IV.4 – The variation of the Privacy coefficient y from 0.1 to 0.9 implies a variation of the trade-off between Utility and Privacy.

	Steps	DTW
Raw data	14387	-
DYSAN	15321 **(+6.49 %)**	12.96
GEN	12817 **(-12.25%)**	14.28
Olympus	23658 **(+64.44%)**	156.03
MSDA	18624 **(+29.45%)**	23.37

TABLE IV.4 – The sanitized signal provided by DYSAN appears to be less distorted and more useful for step detection than other approaches.

Table IV.4 shows that with DYSAN the estimation of the number of steps only suffers from a 7% error compared to the raw data. With the different baselines, the sanitized signal appears to be much more noisy and the step detection is greatly impacted with an overestimate number of the steps of more than 64% for Olympus, more than 29 % for MSDA and more than 12% of errors for GEN. While GEN does not take into account data distortion compared to MSDA, its signal appears to be less noisy with a smaller error rate. This smaller utility loss is balanced by a smaller privacy guarantee as shown in Figure IV.5. Moreover the data distortion improvement of DYSAN compared to MSDA (while ensuring a better privacy) is provided by the dynamic model selection that adapts the sanitization according to the incoming signal at the user level in order to have the best utility-privacy trade-off. This fine-grain sanitization is not possible through a single sanitizing model for all users and whatever the performed activity (*i.e.*, the incoming signal). The method ORF is not considered here because it only extracts features and the signal is not preserved, which prohibits possibility to conduct further analysis.

To evaluate the deformation of the signal, we also report the Dynamic Time Warping (DTW) [39] between the raw and the sanitized data from each baseline (Table IV.4). Based on this alignment algorithm, we can measure the distortion between two temporal signals with a distance metric (in our case an euclidean distance). If this metric has a small value then it means that the two signals are quite similar to each other, which is a sign of a small distortion. The results obtained show that the sanitized data produced by DYSAN is more similar to the raw data compared to other baselines. Similarly to step detection, the sanitizing process of Olympus depicts a large data distortion making further analysis of the signal impossible.

Table IV.5 gives complementary results concerning the similarity analysis of the data

sanitized between the different baselines, with simple quantitative measures. Here the raw measures plus the percentage relative error are given for each baselines. Even if those metrics give few information about the shapes of the signals, we can still observe that Olympus, the only baselines that does not take into account the distortion of the data during training, is the one that has his measures very far from the raw data. For example the standard deviation is almost five times higher than the original data showing a very noisy signal.

	Mean	Std	Skewness	Kurtosis	Energy
Raw	0.81	0.47	1.65	4.81	139.06
DySan	0.68 (-15.9%)	0.77 (+62.9%)	0.40 (-75.7%)	1.28 (-73.5%)	230.87 (+66.0%)
GEN	0.28 (-65.4%)	0.12 (-74.7%)	0.51 (-69.2%)	0.08 (-98.3%)	12.11 (-91.3%)
Olympus	5.40 (+566.4%)	2.52 (+433.1%)	0.61 (-62.8%)	0.29 (-94.0%)	4631.47 (+3230.5%)
MSDA	0.54 (-33.5%)	0.24 (-49.9%)	0.41 (-75.2%)	-0.11 (-102.2%)	51.87 (-62.7%)

TABLE IV.5 – Similarities metric on the raw data and the different baselines. Mean, standard deviation (std), skewness, kurtosis, energy are given in percentage of relative error.

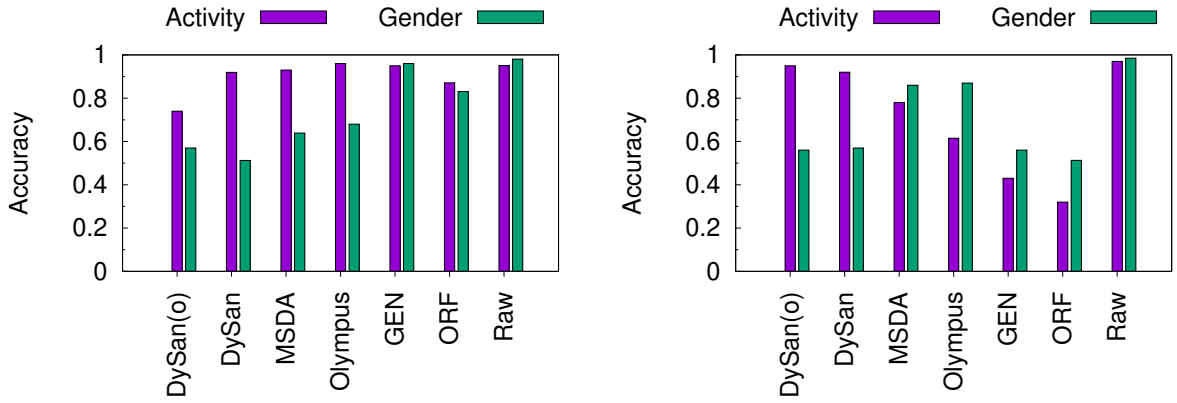
IV.1.4.4 Comparative analysis

We compare DYSAN against baseline approaches (Figure IV.5). Two versions of DYSAN are given to represent, DYSAN where the annotations of the activities are known and the online version, DYSAN(o), where the activities are not given but inferred from the RF classifier. The first version has been added for a more fair comparison to state-of-the-art that does not evaluate models as we suggest.

For MotionSense (Figure IV.5), the privacy improvement of DYSAN occurs at the cost of a slight decrease of utility (gender inference limited to 51% and an activity recognition of 92%). For the online version, which works blindly without annotations, the performance is a little worse, with a gender inference of 57% and accuracy in activity of 75%. This utility mitigation comes from the imperfect accuracy of the RF classifier used in the online phase to select the best sanitized model. Indeed, to dynamically select the sanitizer model, DYSAN needs to estimate the model providing the best utility and privacy trade-off with respect to the considered parameters (Section IV.1.2.3). To achieve this, DYSAN relies on a calibration process to build a RF classifier on the smartphone using the raw data used as a reference to predict the current activity performed by the user. This local RF classifier provides an average accuracy of respectively 96% and 94% on the activity recognition for MotionSense and MobiAct datasets. While these accuracies are high, an activity wrongly predicted by this classifier leads to a selection of the sanitizer model that does not correspond to the best utility and privacy trade-off.

As depicted in Figure IV.5, results for MobiAct show that DYSAN and DYSAN(o) outperform other approaches by limiting the gender inference to 55% and 54% while only reducing the accuracy of activity recognition by respectively 2% and 5% compared to using the raw data. Although GEN and ORF also significantly limit the gender inference, the accuracy of the activity detection is drastically impacted (respectively, 43% and 32%). For MSDA, both utility and privacy are not impacted compared to Olympus and GEN due to the distortion limitations that force the model to avoid large transformations on the data.

The accuracy of the classification is not uniform for all activities. Table IV.6 details the True Positives and False Positives of this classification for DYSAN on MotionSense dataset. This table also reports the percentage of data in the dataset for each activity. We observe that the accuracy of the classification depends on the performed activity. This heterogeneity



(a) MotionSense (b) MobiAct

FIGURE IV.5 – DYSAN provides the best privacy protection compared to state-of-the-art approaches at the cost of a slightly smaller accuracy in term of activity detection.

is a direct result of the unbalanced classes. Specifically, the walking activity has the highest precision which corresponds to the activity with the largest amount of data, while other activities contained less data and depicted lower good predictions. This difference in terms of good prediction between walking and other activities can also be explained by a calibration of the size window adapted for the walk (Section IV.1.1).

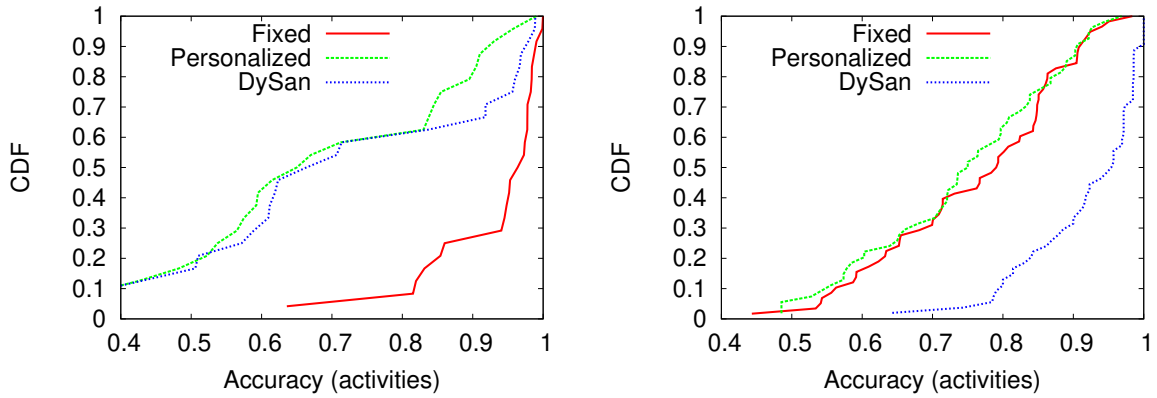
	TP	FP	Precision	Data percentage
Downstairs	221	112	66.4	17.2
Upstairs	223	198	53.0	20.5
Walking	918	74	92.5	44.9
Jogging	216	212	50.5	17.4

TABLE IV.6 – True Positive, False Positive, Precision and percentage of data for each activity of Dysan (MotionSense dataset).

Results also show the performance improvement provided by each baselines approach based on adversarial networks. Specifically, GEN, Olympus and MSDA gradually improve the utility and privacy trade-off. However, our utility analysis (Table IV.4) shows that the sanitized data is very distorted, which harms the possibility to perform signal processing for further analysis. MSDA integrates the data distortion in its loss function, which leads to less distorted data. This feature improves the quality of signal processing but does not significantly improve the trade-off between utility and privacy compared to Olympus (Figure IV.5). By dynamically selecting the best sanitizer model for each window of raw data, DYSAN(o) makes the gender inference close to a random guess while preserving an accurate activity detection.

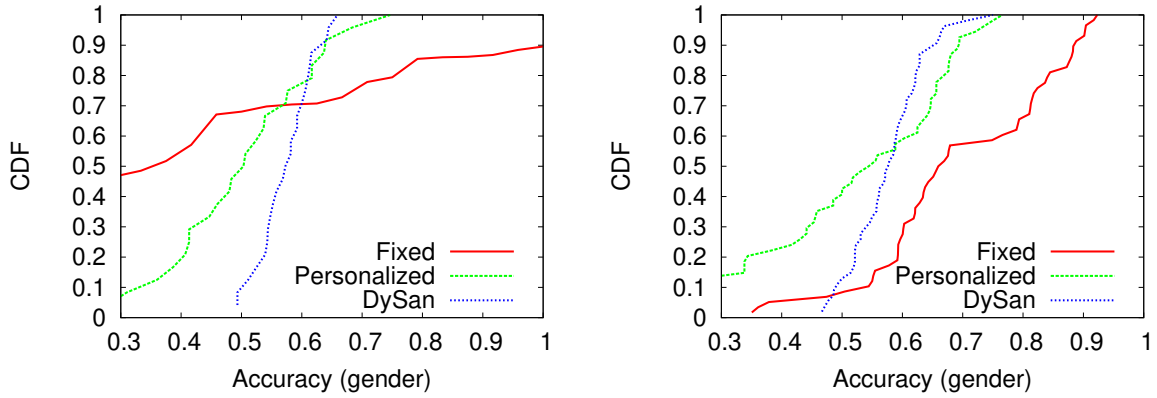
The results of GEN reported in [199] mention an accuracy of 94% for the activity recognition and 64% for the gender inference for MotionSense dataset compared to 95% and 96%, respectively in our experiments. This difference comes from our implementation that exploits two neural networks for each classification task (*i.e.*, for activity recognition and gender inference) versus only one neural network for both classification tasks in the original baseline as explained in Section IV.1.3.2). However, this difference also tends to assume an over adaptation of the underlying neural network to the considered dataset. This over-adaptation is also pointed by the completely different trend for the accuracy provided for

MobiAct compared to MotionSense.



(A) MotionSense (B) MobiAct

FIGURE IV.6 – The dynamic sanitizing model selection of DYSAN significantly improves the activity recognition in case of transfer learning (*i.e.*, MobiAct dataset).



(A) MotionSense (B) MobiAct

FIGURE IV.7 – By dynamically adapting the sanitizing model for each user according to the incoming data, DYSAN greatly improved the protection against gender inference (the distribution of the gender accuracy is more centered around 0.5, which corresponds to a random guess).

IV.1.4.5 Dynamic selection of sanitizing model

During the training phase, DYSAN computes the sanitizer models corresponding to all possible utility and privacy trade-offs by exploring the range of values for the hyperparameters α and λ. We evaluate here the benefit to dynamically adapt the sanitizing model according to the incoming data of each user compared to two static baseline approaches. Firstly, we compute the accuracy for both the gender inference and the activity recognition when the sanitizer model is fixed for all the users. This case represents the behaviors of all comparative baselines where the considered model is the one providing the best performance (*i.e.*, the utility and privacy trade-off) on average for all the users. Secondly, we consider a personalized solution in which the sanitizer model is personalized for each user.

In this case, the sanitizing model is the one which provides the smallest accuracy in terms of gender inference and the best accuracy in terms of activity recognition according to the whole models set for a specific user. This solution provides a sanitizer model personalization but the selected model is static and does not change according to the evolution of the incoming data (and the associated changes in terms of performed activity).

We compare these static solutions against DySAN in which the considered sanitizer model for each user changes according to the incoming data in order to maximize the utility and privacy trade-off over time. Figures IV.6 and IV.7 depict for both datasets the cumulative distribution (*i.e.*, CDF) of the accuracy of the activity recognition and the gender inference respectively, when a fixed, a personalized and a dynamic sanitizing model is considered. Results show that the accuracy in both classification tasks is highly heterogeneous over the population of users. This high heterogeneity reflects the fact that a static model is not well adapted for all users or for all activities performed by the user, thus motivating the need for a dynamic approach.

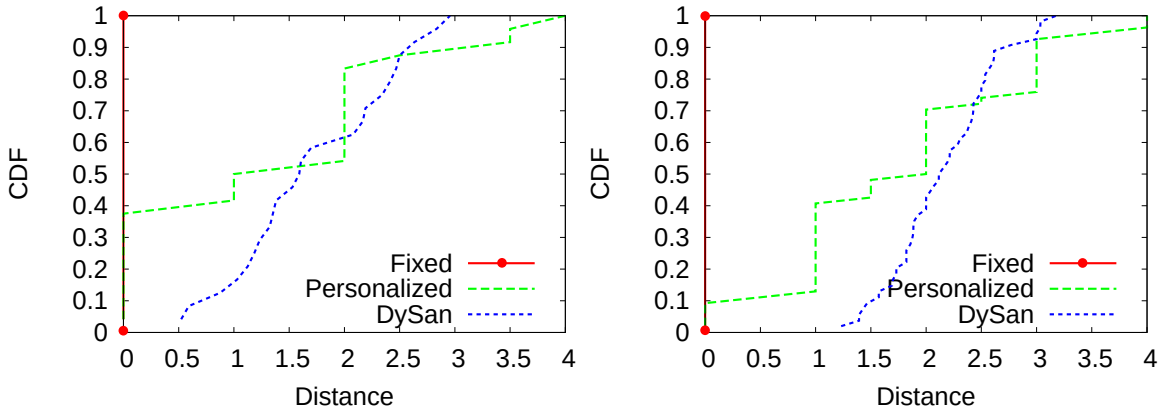
Specifically, results show that dynamically adapting the sanitizing model significantly improves the activity recognition compared to using a static model in case of transfer learning (*i.e.*, MobiAct dataset, Figure IV.6b). For MotionSense dataset (Figure IV.6a), most users benefit from an important accuracy with a static model fixed for all users. This result can be explained by the fact that the sanitizing models have been learned with the same users, leading to a learning of the motion characteristics of all the considered users.

For the gender inference, the objective of the sanitizer is to provide an accuracy around 0.5 which corresponds to a random guess for all users. However, results depicted in Figure IV.7 clearly show that a fixed model for all users fails to protect against gender inference. Indeed, the distribution reports a wide range of accuracy over the users where it is possible to infer the gender with 80% of confidence for 60% and 20% of the users for MobiAct and MotionSense dataset, respectively. Adopting a personalized sanitizer model for each user decreases the accuracy of the gender prediction compared to a fixed model for all users but the distribution of the accuracy is still large (from 0.3 to 0.75 for MotionSense and from 0.3 to 0.8 for MobiAct). By dynamically adapting the sanitizing model according to the incoming data, DySAN greatly improves the protection against gender inference compared to using a fixed model with a sharper distribution centered around 0.5.

These results also show the capacity of DySAN to transfer the learning performed on MotionSense to MobiAct (an activity recognition accuracy around 92% on average for a gender accuracy around 57%). For comparison, we evaluated the transfer learning of GEN using the original sanitizing model learned on MotionSense (and publicly available) to the MobiAct dataset. In this case of transfer learning GEN provides an accuracy in terms of activity recognition and gender detection around 43% and 56%, respectively. This result shows the limited capacity of GEN to transfer learning from MotionSense to another dataset assuming an over adaptation of the underlying neural network and parameters to the considered dataset.

To go further, we evaluate the variation of the sanitizer model selection of DySAN compared to static approaches using either one model fixed for all users or one personalized model for each user. To achieve that, we measure the distance between the hyperparameters α and λ corresponding to the best privacy and utility trade-off on average for all users (*i.e.*, the model fixed for all users) and the model selected for each user (*i.e.*, a personalized model) or according to the incoming data (*i.e.*, the model dynamically selected by DySAN). Figure IV.8 reports the distribution of this distance for both datasets. Results show that almost 40% of the users of MotionSense dataset have a personalized sanitized model which corresponds to the model providing the best trade-off on average for all users. In addition,

for both datasets, results show a large variability in terms of distance over all users highlighting the necessity to provide a variety of models to adapt the sanitization.



(A) MotionSense (B) MobiAct

FIGURE IV.8 – DYSAN provides a large variability in terms of distance over all users highlighting the necessity to provide a variety of models to adapt the sanitization.

To complete this analysis, we also counted the number of different models used by DYSAN for each user. Figure IV.9 depicted for both datasets the distribution of the percentage of all possible sanitized models (36 in our experiment as presented Section IV.1.3.4) selected by DYSAN for each user. Results show a large range of different models selected ranging from 20% to 50%. This result shows that DYSAN successfully adapts the sanitization according to the evolution of the incoming data.

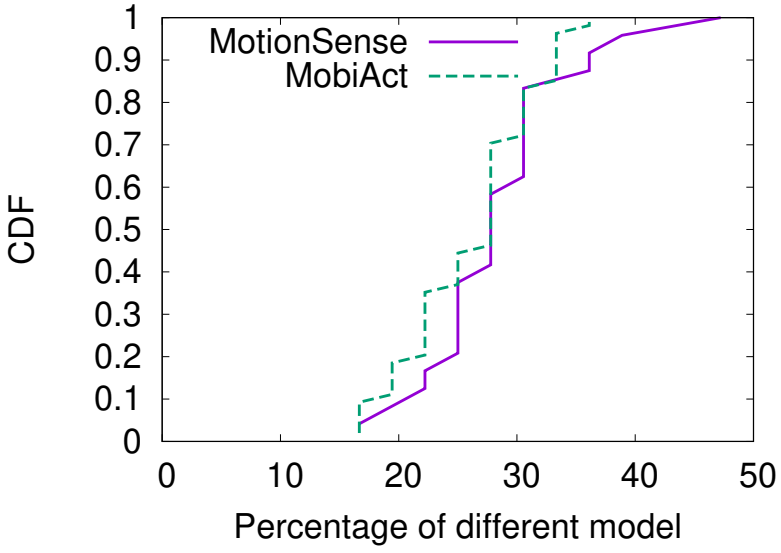


FIGURE IV.9 – The data of each user is sanitized with a wide variety of models (from 20% to 50% of all the models) showing that DYSAN successfully adapts the sanitization according to the evolution of the incoming data.

We also quantify the possibility to use the set of selected models as a fingerprint to identify each user in Appendix IV.1.4.7.

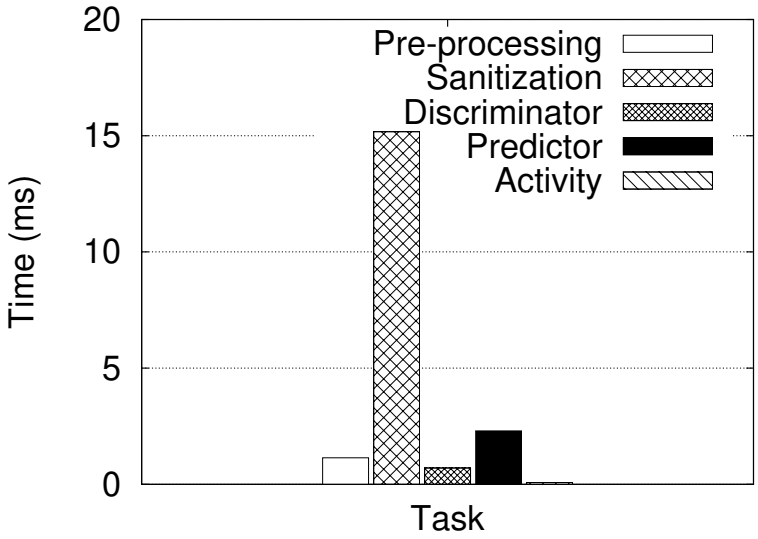


FIGURE IV.10 – The limited cpu overhead of the sanitation of DYSAN is compatible to real-time processing on smartphone.

IV.1.4.6 Performance as measured on devices

We now evaluate the cost of operating DYSAN on a smartphone. DYSAN protects the sensitive attribute while ensuring an accurate activity recognition and minimal data distortion. However, applying the sanitizing at run time on the mobile introduces an overhead. We do not consider the overhead of the learning as it is a one time operation. DYSAN evaluates multiple sanitizing models (*i.e.*, according to each α and λ hyperparameter explored) before selecting the one that produces the best compromise between utility and privacy. Consequently, the overhead associated with the sanitizing of raw data depends on the number of considered models.

Figure IV.10 describes the time (ms) spent by a Xiaomi Redmi Note 7 (equipped with a Qualcom Snapdragon 660 and 3 GB of memory running a java application using Pytorch 1.6) on each task associated to a single sanitizing model of a window of incoming data (*i.e.*, 2.5 seconds of data). Specifically, these tasks include the pre-processing of signals, the sanitizing of raw data, the evaluation of the privacy and the utility on the sanitized data respectively by the discriminator and the predictor, and the classification of the activity performed by the user from the raw data. Excepting the pre-processing, which is performed only once for a window of data, the other tasks have to be repeated for each explored sanitizing model. Results show that applying a sanitizing model once spends most of the time while all operations require 19 ms. Considering 20 or 36 sanitizing models increases this time to 366 ms and 658 ms, respectively. Although this processing is compatible with real-time processing (*i.e.*, data processed after each data window), the number of models deployed on the smartphone should be chosen to limit the overload. The number of considered sanitizer models deployed on the smartphone has also an impact on the storage space requirement. On average, the size of a single model is around 15 MB. Considering 36 models results in 540 MB which is not a limitation with regards to the storage capacities of current smartphones.

We also evaluate the impact of the considered number of sanitizing models. Considering less sanitizing models leads to cover less hyperparameter values and thus limiting the achievable utility and privacy trade-off. Consequently, a degradation of the accuracy for both the activity detection and the gender interference is observed. Table IV.7 presents the

	Activity accuracy (%)	Gender accuracy (%)
36 models	92	57
20 models	89	59
16 models	88	63
8 models	86	66

TABLE IV.7 – Reducing the number of sanitizing models available for the selection decreases the accuracy in activity recognition while increasing the accuracy in gender inference.

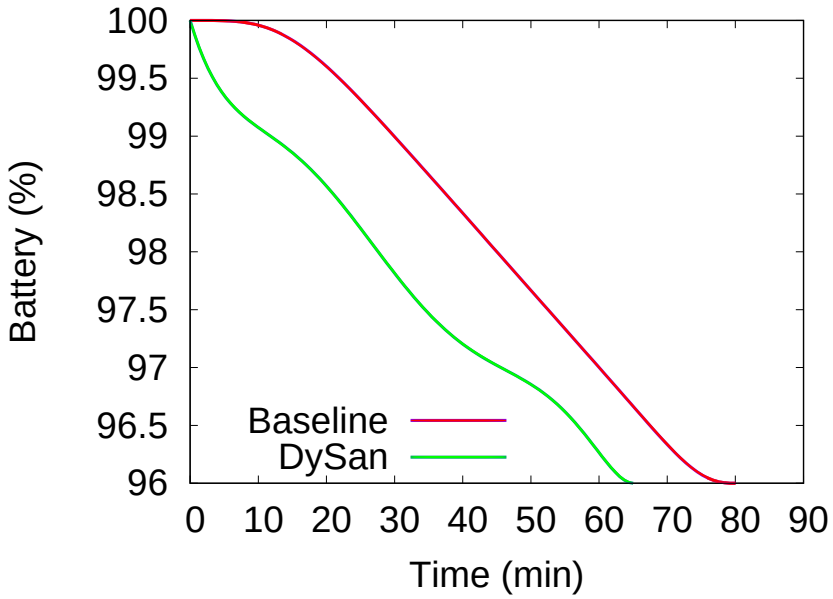


FIGURE IV.11 – The impact of DYSAN on energy consumption is limited (1% less battery after 1 hour).

performances obtained with different numbers of sanitizing models available for the selection. Results show that from 36 to 20 sanitizing models, the accuracy in activity recognition decreases by only 3% and increases by 2% the gender inference.

Finally, we evaluate the impact of running DYSAN on the energy consumption on the smartphone. Figure IV.11 reports the decrease in the battery charge over time for a baseline where no operation is performed on the smartphone, and for a real-time processing of DYSAN (*i.e.*, after each window of raw data, and exploring 36 sanitizing models before to select the best one). In both cases, the screen remained on during the experiment. Results show that DYSAN consumed 1% more battery after 1 hour, which stays a reasonable energy consumption.

IV.1.4.7 Information leakage in model selection

As DYSAN dynamically selects the sanitizing model to use for each window of incoming data, the set of selected models could be leveraged to identify each user. Indeed, this set of sanitizing models chosen by a user could act as a unique fingerprint. To evaluate this potential information leakage, we quantify the uniqueness following the methodology presented in [215]. More precisely, the uniqueness for each user is estimated as the percentage of 100 random sets of p selected sanitizing models that are unique. Figure IV.12 reports for MobiAct dataset the distribution of the uniqueness with p (*i.e.*, the size of fingerprint) from 1 to 5 and with different number of sanitizing models available for the selection. As

expected, results show that the larger the fingerprint, the more unique the behaviour of a user becomes. However, at least 5 models are needed to have a strong confidence (around 80% of uniqueness) when 36 sanitizing models are exploited. To reduce this uniqueness, a lower number of sanitizing models (*i.e.*, through the hyperparameters values explored in the training phase) should be proposed. Indeed, less choice for model selection leads to having more users who share common models. Results show that exploiting less available sanitizing models reduces the uniqueness.

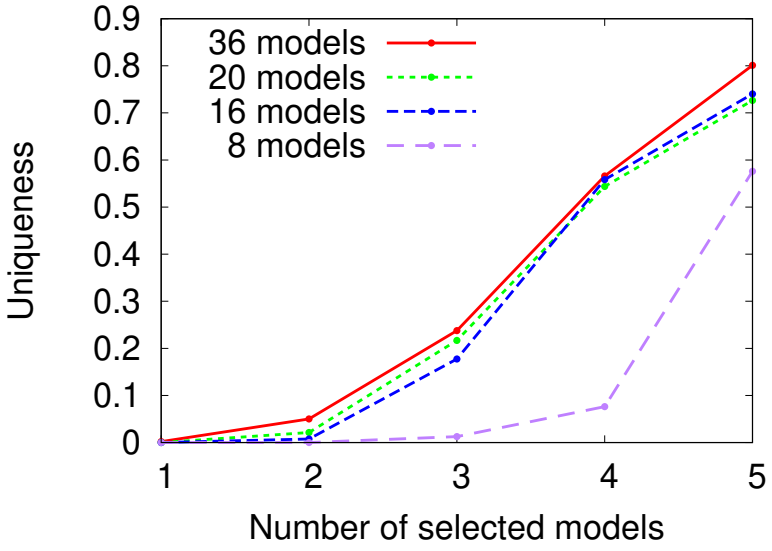


FIGURE IV.12 – The uniqueness of the selected models remains low for fingerprints with less than 5 models, and depends on the number of available sanitizing models for the selection.

Information leakage in model selection leading to user re- identification is only possible if the adversary is able to characterize each selected sanitizing model from the sanitized data. In this case, the adversary could maintain a fingerprint per user to conduct its re-identification attack. To evaluate this capability, we measure the level of distortion using the Dynamic Time Warping of the sanitized data for each sanitizing model. Over all sanitizing models, our results show a very low standard deviation of the DTW. This low value indicates a small difference in terms of distortion when different sanitizing models are exploited, thus making it difficult for an adversary to identify the selected model from the sanitized data. This re-identification attack consequently seems difficult to achieve.

IV.1.5 Conclusion

Globally this privacy-preserving framework by sanitizing motion sensor data presents several benefits compared to the previous contributions.

By adapting the architectures of the different neural networks and their corresponding objective functions, the method can be used for different privacy issues. The method is not specific to only one privacy objective such as the gender or the identity as in the previous chapter.

Moreover, by comparing with the approaches detailed in the previous chapter, here one of the objectives was to limit as much as possible the level of transformation of the raw data. DYSAN preserves as much as possible the useful information for activity recognition and other estimators of physical activity monitoring. Results show that DYSAN drastically

reduces the risk of gender inference without impacting the ability to detect the activity or to monitor the number of steps.

We also showed that the dynamic sanitizing model selection of DySAN could provide a personalized privacy protection by adapting the protection to each user over time according to the evolution of the incoming data. This method is particularly effective in a transfer learning case unlike the other baselines that have a unique sanitizing model for all the users. Moreover, the overhead introduced on the smartphone to sanitize the data is compatible with real-time processing while keeping a reasonable energy consumption. By comparing our approach with existing approaches, we demonstrated that DySAN provides better control over privacy-utility trade-off.

Concerning limitations, the framework is based on the hypothesis that the dataset to be sanitized is balanced, which means that there is no direct correlation between the utility classification (*i.e.*, activity recognition) and the sensitive information targeted. If this assumption is not respected for a specific application, the sanitization would probably not be effective. In addition, the framework specifically focuses on one sensitive information at one time, the privacy objective of DySAN only focuses on gender and no other sensitive attribute.

Sharing sanitizing data with an application server remains unsafe concerning any other sensitive attribute inference not concerned by the sanitization. Another solution consists in using a distributed framework. Instead of sending the data to the application server, FL propose to keep the user's data locally and only share ML models. However, sharing models can also raise other privacy issues which will be developed in Section IV.2).

IV.2 Privacy Assessment of FL using Personalized Layers

This section is a short evaluation of a specific FL method that (as the previous contributions) minimize the quantity of information communicated to the cloud and provide at the same time a personalized response to each user's data.

FL involves combining ML models from distributed partitions of data and Federated Averaging [206] is the leading optimization method in case of data sharing a similar distribution. However, with practical deployment scenarios such as the presence of non independent and identically distributed data, the performance of Federated Averaging can be severely degraded especially with atypical users.

At the same time, while FL improves privacy by reducing the exposition of the personal data, it remains vulnerable to threats such as poisoning attacks, membership inference attacks, etc. To mitigate the risks, several approaches have been proposed from using Differential Privacy locally at user level or server level [224], Homomorphic Encryption (HE) and Secure Multiparty Computation (SMC) [90] (See Section II.3).

The development of FL highlighted other challenges [189] such as the heterogeneity of data across user devices leading to a degraded accuracy for less represented users. To overcome this limitation, [25] studied an FL scheme using personalization layers. In this scheme, the local model on each participant is composed of *lower layers* (capturing coarse grain information) trained following classical FL learning round, and *upper and personalization* layers (capturing fine grain information) trained locally and which stay private on the device and not exchanged with the server (Figure IV.13). This scheme is known to improve the accuracy of the model in presence of heterogeneous data across users.

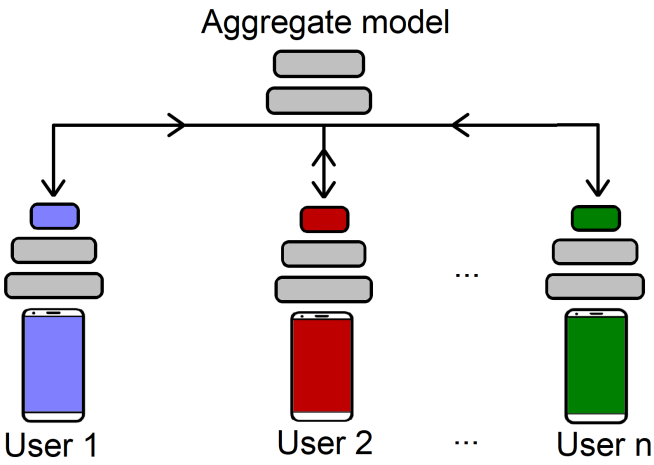


FIGURE IV.13 – Personalized FL approach: only the upper layers (colored in grey) are shared with the server while the personalization layers are kept private on the device.

However, the privacy impact of sharing only a sub part of the model has never been measured. We quantify in this section the utility and privacy of a FL scheme using private personalized layers [25]. To assess privacy leakage, we consider both an attribute and a membership inference attack. Evaluations have been conducted using two datasets of motion sensor data collecting in real-life conditions.

IV.2.1 Evaluation

We exhaustively evaluate the utility and privacy of a FL scheme using private personalized layers in the context of activity recognition (details of the methodology are given

Section IV.2.1.1). In this section, we show that personalized layers improves the utility (Section IV.2.1.2) and privacy (evaluated through attribute inference Section IV.2.1.3 and membership inference Section IV.2.1.4) compared to both a vanilla FL and a defense scheme using local differential privacy.

IV.2.1.1 Experimental setting

System model: We consider a FL scheme using SGD addressing activity recognition. The learning model is based on 2 convolutional layers, and 3 fully connected layers. Only the 2 lower convolutional layers are exchanged with the server which aggregates and disseminates model updates to devices, the 2 upper fully connected layers stay private on the device and are personalized with the user data. The devices of users are considered as trusted but it is not the case of the server which is considered as an adversary trying to infer personal information of participants from their model updates.

Datasets: Two real-life condition datasets are used for the evaluation. They are both publicly available and heavily used in the literature. These datasets come from the extraction of motion sensor data during gait activities (*i.e.*, based on step patterns) of different subjects.

— **MotionSense** [198] contains motion data captured from an accelerometer (*i.e.*, acceleration and gravity) and gyroscope of an iPhone 6s kept in the front pocket at a frequency rate of 50Hz. Overall, six activities (*i.e.*, walking, jogging, going upstairs, going downstairs, sitting and standing) have been made by 24 users during 15 trials in the same conditions and environment.

— **MobiAct** [317] records the motion data from 58 subjects during more than 2500 trials, all captured with a smartphone also in the front pocket. This dataset includes signals recorded from the accelerometer and gyroscope of a Samsung Galaxy S3 smartphone. Nine different activities of daily living are performed by the users. We only used the trials corresponding to the same activities as MotionSense in order to do the evaluation with the exact same settings.

Both datasets contain an equal number of men and women, and each activity is performed according to the same conditions by all subjects. However, the walking activity is more represented than the others. For each user, we also have access to physical information such as the gender, weight, height and age.

Baselines: We considered two baseline approaches to compare FL scheme using private personalized layers (**FedPer**) [25]:

— **Standard FL (Vanilla)** [206] This is the most common FL scheme using SGD training on the device and average aggregation of all models at each learning round on the central server.

— **Local Differential Privacy (LDP)** [224] We consider an implementation based on an introduction of noise following a Gaussian distribution ($\mathcal{N}(0,0.01)$) to the model updates computed through a classical learning phase (*i.e.*, Vanilla).

Evaluation metrics: We evaluated FedPer and the different baselines along both utility and privacy metrics.

— **Utility:** To measure the utility, we considered the accuracy of the predicted activity. More precisely, we produce a confusion matrix based on the output of the classifier and measure the number of correct predictions made by this classifier over all predictions made. The value of the accuracy ranges from 0 to 1, in which 1 corresponds to perfect accuracy.

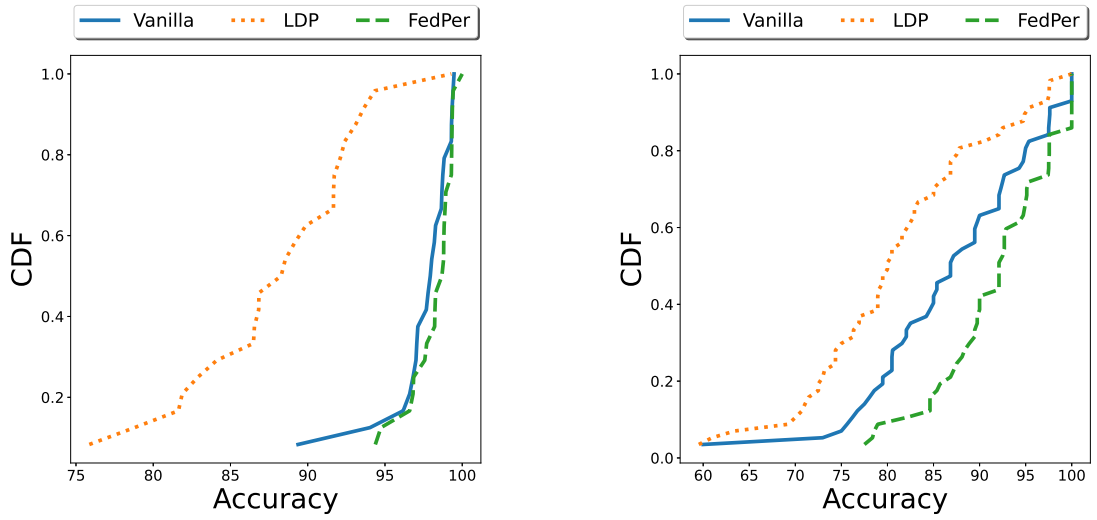
— **Privacy:** To assess the level of privacy, we rely on the accuracy of both the inference of sensitive attributes and the inference to be a member of the training set. These inference attacks implement the solution proposed by [109] which leverages an invariant

permutation representation of nodes at each layer to classify model updates received by the server through a RF of 1000 trees with a maximum depth of 10. We consider the gender and the Body Mass Index (BMI) of the users as sensitive attributes. The BMI is a value defined by the weight of the user divided by the square of her height. This value allows to categorize a person as underweight, normal weight, overweight or obese. In our case we only focus on a binary classification: overweight (BMI $>$ 25) or not (BMI $<$ 25) for the sake of class balance. For the membership inference, the accuracy refers to the percentage of correct prediction (that a participant has been involved in the training of the model) over all predictions made. In both attacks, an accuracy of 0.5 corresponds to a random guess as our dataset is balanced.

Implementation details: For each experiment, we run 10 times of 5-fold cross validation where each fold is tested based on the training of the other four. We considered 200 learning rounds and an early stopping that stops the learning process if the average test loss of the aggregated model sent locally on the user data does not decrease during 30 learning rounds. During each learning round, the training with SGD is done locally at the user's level during 10 epochs. A constant learning rate is used with $\eta = 0.001$ for all the users.

IV.2.1.2 Utility Evaluation

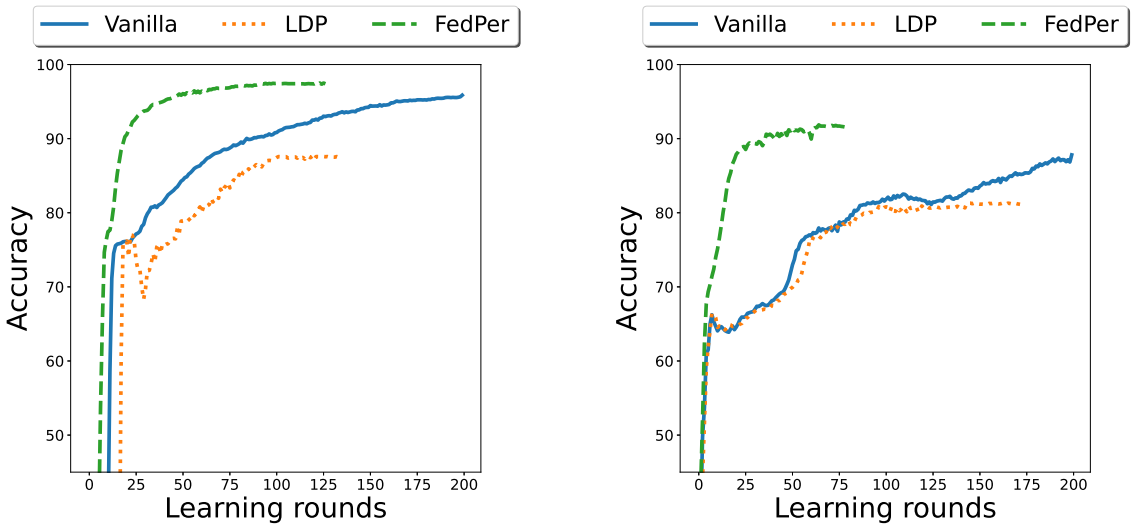
We measure the accuracy of the activity detection of FedPer and the baselines. Figure IV.14 reports the Cumulative Distribution Function (CDF) of this accuracy over the population of users for MotionSense and MobiAct dataset. First, results show that the local adaptation of FedPer on the upper layers slightly increases the accuracy compared to the Vanilla approach (from 1% to 7% of increase on average for MotionSense and MobiAct, respectively). Second, results show that LDP baseline degrades significantly the accuracy for both datasets (10% on average of MotionSense and 6% on average for MobiAct). Indeed, by introducing noise, the convergence of the model is greatly degraded leading to a loss of prediction for all users. This result comforts previous results [195, 343].



(a) MotionSense (b) MobiAct

FIGURE IV.14 – By personalizing upper layers of the model, FedPer sightly increases the accuracy of the activity prediction compared to a FL vanilla approach; local differential privacy, in turn, greatly degrades the accuracy.

We also measured the convergence speed of the learning. Figure IV.15 depicts the accuracy of the activity detection as a function of learning rounds for FedPer and the Vanilla approach. Results show that FedPer drastically speeds up the convergence. For instance, FedPer achieves 90% of accuracy after 12 learning rounds on MotionSense where the Vanilla approach achieves the same accuracy after 100 learning rounds. For MobiAct, FedPer achieves 90% of accuracy after 35 learning rounds where the Vanilla approach only reaches 86% of accuracy after 200 learning rounds. By using its personalized layers at each learning round instead of starting the learning from the aggregate model sent by the server, the accuracy increases faster. For LDP, we can observe that the noise introduced prevents the model from converging to an optimal model but a sub-optimal one.



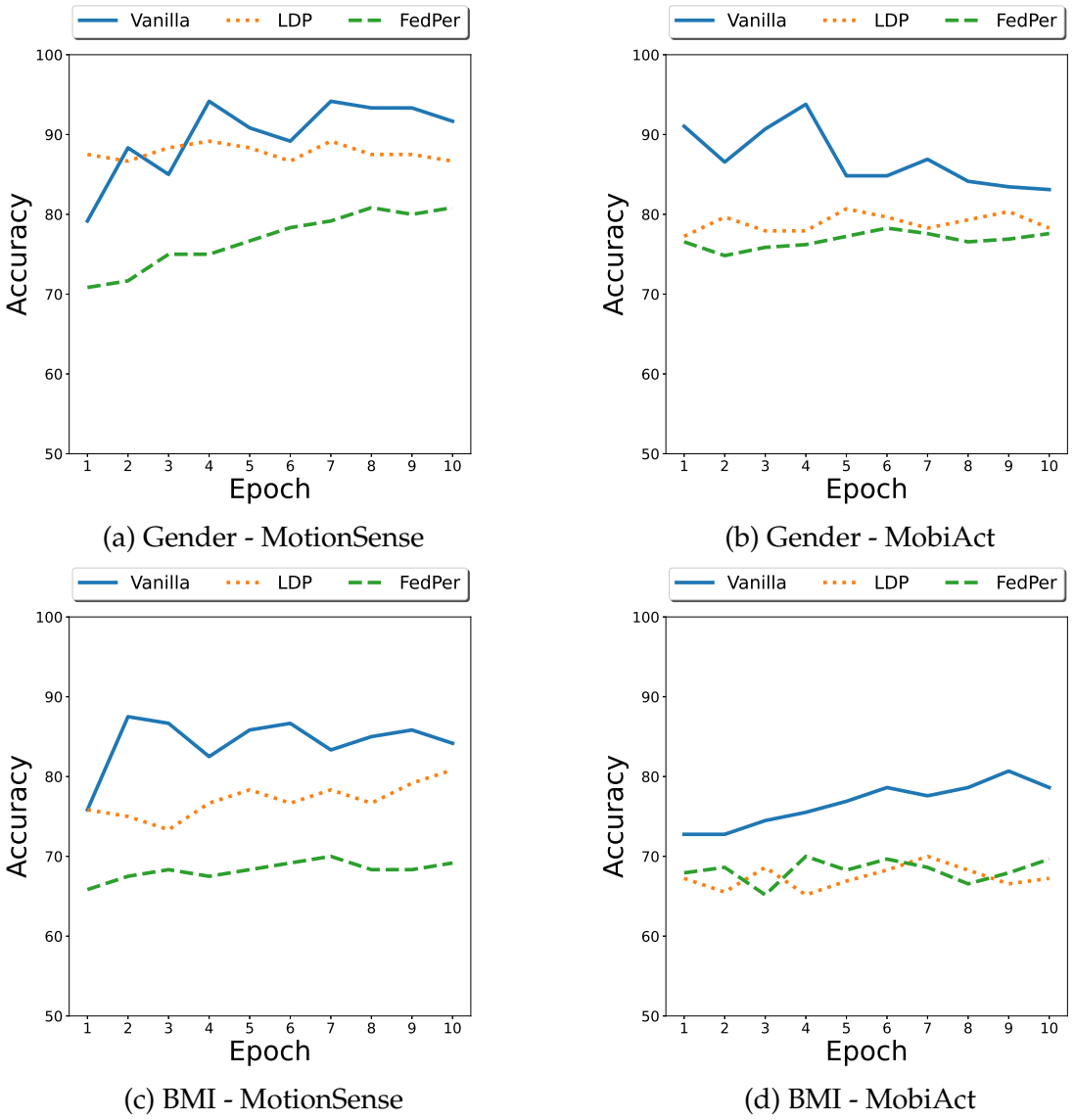
(a) MotionSense (b) MobiAct

FIGURE IV.15 – By using personalized layers instead of aggregated information, the learning is drastically speeds up.

IV.2.1.3 Privacy evaluation through attribute inference

We conducted an attribute inference attack to infer the gender and the BMI of users from their model updates sent to the server. In this attack, we use half of the participants to train their local model on 80% of their data. Once all the models are sent to the server, only the models from one class of the targeted attribute are aggregated (in our case, models from women for gender inference, and models from overweight users for BMI inference). This malicious aggregated model is representative of one class of the attribute inference targeted. Then the malicious server sends back this aggregated model to the remaining users that did not participate in the first aggregation to fine-tune locally on the remaining 20% of their data (e.g., training from a model aggregating model updates from women) before returning the update to the server. The adversary then trains an RF classifier on these model updates to infer the sensitive attribute. This training exploits 80% of all the updates and the testing is done on the remaining 20%, with cross validation.

Figure IV.16 evaluates, for both datasets, the accuracy of these both sensitive attribute inferences over the epochs of local learning. Firstly, results show that without any protection (*i.e.*, the Vanilla approach), all sensitive attributes can be inferred with high accuracy for both datasets (e.g., around 90% of accuracy for the gender on MotionSense). FedPer reduces



(a) Gender - MotionSense

(b) Gender - MobiAct

(c) BMI - MotionSense

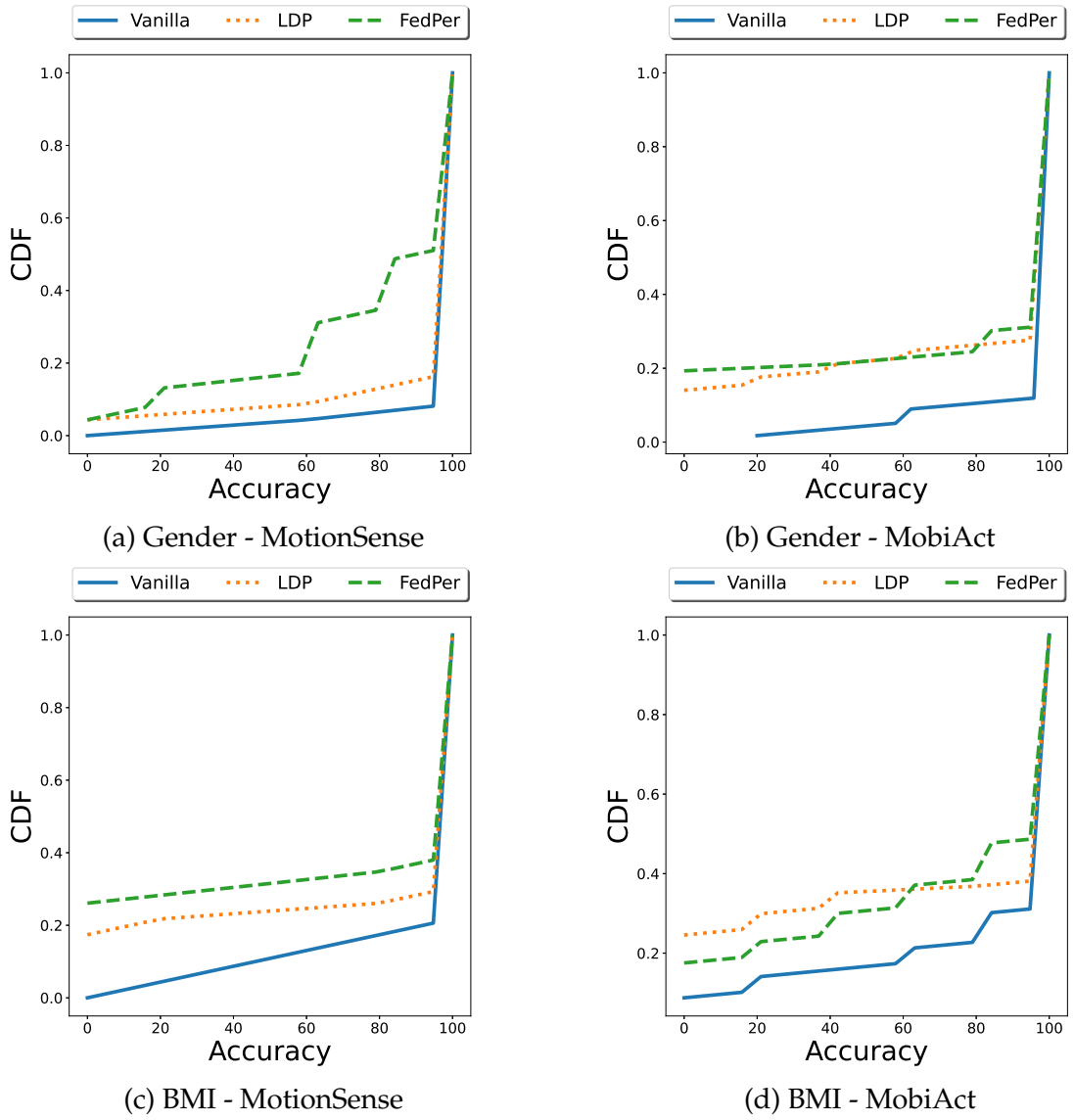
(d) BMI - MobiAct

FIGURE IV.16 – The increase of number of learning epochs per user increases the accuracy of the attack on both sensitive attributes.

this accuracy between 10% and 20% according to the dataset and the sensitive attribute. Results also show that FedPer better protects users against inference attack compared to LDP regardless of dataset and sensitive attributes (from 5% to 10% of accuracy loss for MotionSense).

Secondly, results show that the inference accuracy tends to increase over the epochs for all approaches. This is explained by the fact that attribute inference attack is closely related to overfitting [341], the more the model learns on user's data, the more it adjusts the parameters to data structure and the more it may incorporate sensitive information.

Figure IV.17 reports the CDF of the inference accuracy over the participants. Results show that while each attribute can be inferred with high accuracy for a large part of the users, this accuracy drops for few percent of users. FedPer and LDP increase the percentage of users with a small inference accuracy.



(a) Gender - MotionSense

(b) Gender - MobiAct

(c) BMI - MotionSense

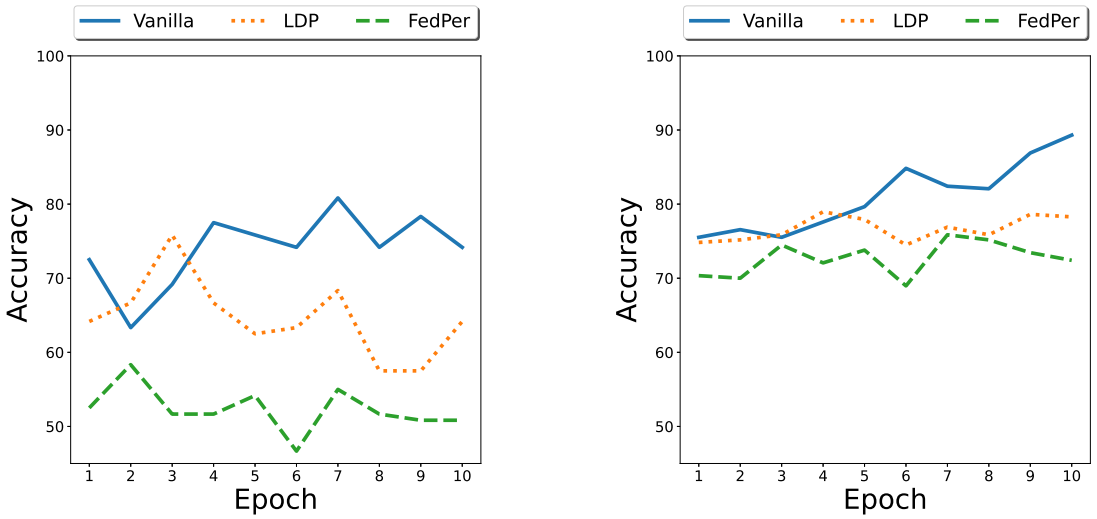
(d) BMI - MobiAct

FIGURE IV.17 – FedPer and LDP increase the number of users with a small inference accuracy.

IV.2.1.4 Privacy evaluation through membership inference

Lastly, we conduct a membership inference attack to evaluate privacy. In this attack, 50% of the users follow a normal FL learning round with 80% of their data. The models are sent to the server which disseminates back the aggregated model to all the users. All of them fine-tune the aggregated model on their remaining 20% of data. The server then trains a RF to classify membership from model updates for all users (using 80% of all these updates for the training and 20% for the testing with cross validation as described in section IV.2.1.1).

Figure IV.18 depicts the accuracy of this inference attack for both datasets and for all approaches. Similarly to the attribute inference attack, results show that the membership inference attack is more efficient on the Vanilla approach. FedPer provides the best protection compared to LDP (20% on average for MotionSense dataset). Interesting enough, FedPer depicts an accuracy close to 50% which corresponds to a random guess (if the data of a specific user has been used to train the model) for MotionSense dataset.



(a) MotionSense (b) MobiAct

FIGURE IV.18 – FedPer and LDP significantly decrease the accuracy of the membership inference attack compare to Vanilla method

IV.2.2 Conclusion

In summary, this chapter focused on the protection against sensitive attribute inference by firstly using a framework based on deep neural networks to automatically sanitize data and obfuscate a targeted attribute while keeping the sanitized data as close as possible to the raw data.

Secondly, we experimentally quantified the utility and privacy trade-off of FL using private personalized layers proposed by [25] in a context of activity recognition. We consider both an attribute and a membership inference attack to measure privacy leakage. Results show that using private personalized layers provides a better utility and privacy trade-off compared to a FL vanilla approach and a defense scheme using local differential privacy. Results show that FL with personalized layers speeds up the convergence compared to vanilla FL and slightly increases the activity accuracy between 1% and 5%, while decreasing the gender and the overweight inference between 10% and 20% and 15% on average for membership inference. This utility and privacy trade-off is better than a defense scheme using local differential privacy which decreases the inference of the gender and the overweight up to 12% but at the cost of the activity accuracy which reduces up to 10%.

These results tend to show that minimizing the information exchanged with the server is an interesting avenue for both personalizing the model (*i.e.*, improving accuracy) while limiting potential inferences (*i.e.*, improving privacy).

Chapter V

Conclusions and perspectives

This chapter summarises the main results developed during the three years of my thesis, and emphasises the associated research perspectives that appear to be in line with the contributions presented in this manuscript.

In this thesis, we investigated different aspects of privacy through ML utilized in the context of gait monitoring. Each contribution developed has two correlated objectives: 1) Giving privacy protection that can preserve motion sensor data from exposure to sensitive information theft (*i.e.*, re-identification, gender or BMI inference). 2) Maintaining the usefulness of this protected data for a gait monitoring application such as activity recognition. Each contribution aimed at giving the most effective trade-off between these two objectives and also intended to overcome some limitations given by the previous contribution.

V.1 Overview of contributions and perspectives

A review on the contribution of ML in the validation of commercial wearable sensors for gait monitoring in patients. On the basis of the review in the chapter II.2, we could observe that wearable sensors validation with ML takes an increasingly important place in the literature, with a number of studies having gradually increased since 2010. In these studies, a significant part of the validation was based on traditional statistical approaches (75%) with a smaller contribution of ML-based approaches (25%). This scoping review highlights the current state of the ability of commercial sensors to enhance traditional methods of gait assessment and identified different recommendations for data acquisition, collection, processing and validation, in order to use sensors in a good way. As long as the data collected are numerous, annotated, and representative, ML is the best approach to interpret and extract valuable information in multi-dimensional data space.

Anonymization framework through minimization. In the chapter III.1, we presented a framework which relies on a ML technique to efficiently recognise the user activity pattern, useful for personal healthcare monitoring, while limiting the risk of re-identification of users from biometric patterns that characterizes each individual. To achieve that, we firstly deeply analysed different features extraction schemes in both temporal and frequency domains. We highlighted that features in the temporal domain are useful to discriminate user activity while features in the frequency domain lead to distinguish the user identity. On the basis of this observation, we secondly designed a novel protection mechanism that processes the raw signal on the user's smartphone and transfers to the application server only the relevant features unlinked to the identity of the users. In addition, a generalisation-based approach is also applied on features in the frequency domain before being transmitted to the server in order to limit the risk of re-identification. We extensively evaluated our framework with a reference dataset: results show an accurate activity recognition (87%) while limiting the re-identification rate (33%). This represents a slight decrease of utility (9%) against a large privacy improvement (53%) compared to state-of-the-art baselines.

However, by minimizing information with a few features, the possibility to have a better trade-off than the one obtained in this contribution, is very reduced. The accuracy of 53% of

re-identification shows that the features used for activity recognition contains information that allow a model to perform part of the re-identification. This limitation can be addressed in two ways. Either applying other statistical transforms on the remaining features following a similar approach or searching for another representation of the data in order to apply a finer minimization method that could better target the re-identification task. The first way presents the risk of over-damaging the features so that the trade-off would not be so beneficial, so that the second way was chosen. The next contribution was then a direct answer to the limitations raised by the first contribution. By still considering that the temporal and frequency domains contain activity and identity information, the objective was to find a data representation that includes both domains in order to apply anonymization methods that could perform a better trade-off. Finally, it would be useful to do a deployement on real-life cases on hospital to benefit from the feedback of clinicians.

Motion sensor data anonymization by time-frequency filtering. The chapter III.2 proposed a novel anonymization framework which consists of a two-step process. First, acceleration signals are encoded in the time-frequency domain by three different linear transforms: the Short Time Fourier transform, the Stockwell and Optimized Stockwell transform. Second, we proposed a method to anonymize the acceleration signals by filtering in the time-frequency domain. Finally, we evaluated our approach for the three different linear transforms with a neural network classifier by comparing the performances for activity versus identity recognition. We extensively studied the validity of our framework with a reference dataset: we determined that the optimized S-transform gives the best utility-privacy trade-off by filtering its TF coefficients at 70%. Results show an accurate activity recognition (85%) while limiting the re-identification rate (32%).

Although this contribution presents interesting results similar to the previous contribution by a few percent, it depicts several limitations that can be overcome with different extensions which could lead to the improvement of the trade-off.

Firstly, only accelerometer signal was considered and using gyroscope signal could improve the utility-privacy trade-off. Rouget and al. [262] proposed a direct extension of this method by including gyroscope TF images. Different scenarios were covered by using different combinations of input data with the x, y, and z axis of the accelerometer and gyroscope, and also by using two configurations of the CNN with the early and late fusion strategies. The first one consists in combining images from the 3 axes at the entry of the network and the second one consists in using three independent convolutional branches to process each input independently.

Secondly, the anonymizing method consisting in filtering high coefficients of the spectograms to remove user's information to prevent re-identification was intentionally naive and allows us to show that the framework provides promising results. Rouget and al. [262] also extended this method by considering STFT and a recent link made by Flandrin al. [100] between STFT and the distribution of zeros in this spectrogram. Moreover, Bardenet and al. [34] explicitly characterize the statistical distribution of the zeros by showing that the zeros of the STFT correspond to the zeros of the Bargmann transform which also correspond to the zeros of Gaussian analytic functions (GAFs) (see mathematical details and properties of these connections in [34]), so that they can establish a precise meaning between the zeros of the spectrogram and white noise signature. In this sense, the distribution of zeros can provide information on the presence of noise or signal for the development of filtering schemes. The intuition behind this idea is that the presence of a signal will modify the distribution of the zeros in the time-frequency domain and mark this distribution by the signal signature, as we can observe by comparing the STFT spectrogram of a white noise and a signal in Figure V.1.

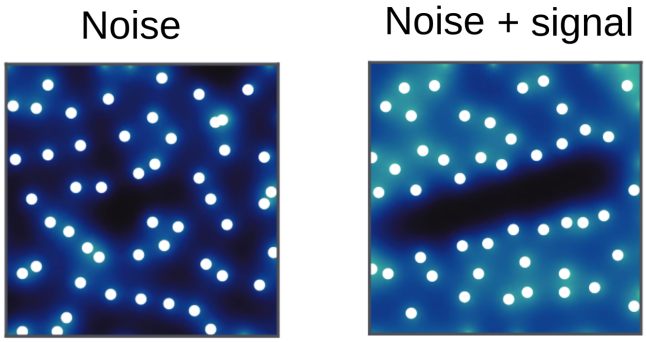


FIGURE V.1 – Representation of the zeros of STFT transform for a white noise and linear chirp signal with a white noise. The signal reveals specific patterns in the zeros while random zeros pattern is associated with the white noise. *Illustration reproduced from Bardenet and al.* [34]

After detecting the zeros in the STFT representation, a graph is created with the zeros localisation as nodes thanks to the Delauney triangulation method [100] (see Figure V.2). Based on this graph, numerous features are calculated on the spectrograms of each axis of each sensor type. Some global features represent statistics of the distribution of zeros, while other local features focused on the zero itself, by calculating its intensity, its coordinates in the image, and the Haralick features [131] that investigate the patterns in the region surrounding the zero. Around 1700 features are then calculated and used for activity and identity classification. Finally, a selection of the features used for activity recognition make it possible to improve the utility-privacy trade-off.

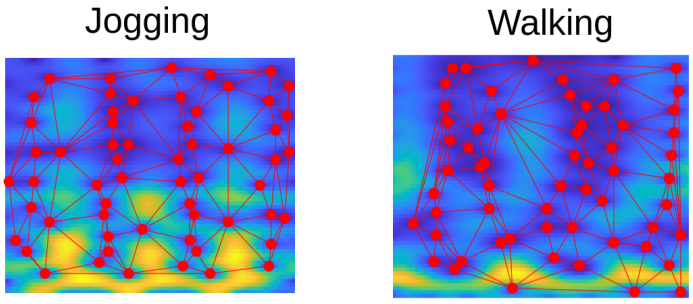


FIGURE V.2 – Examples of STFT representations superposed with the associated graph formed the zeros of the STFTs for different activities (walking and jogging). *Illustration reproduced from Rouget and al.* [262]

This method is specific to anonymization as it is based on the fact that temporal and frequency features are each related to activity and identity information. This framework then could not be applied to other sensitive attributes.

The output data is characterized by a set of features that can only be used for activity recognition. The transformation representation is too far from the raw data that any other observation useful for a clinician could hardly be done. For example, calculating the number of steps, detecting fall, evaluating instabilities in the gait, etc. could not be done with the remaining features. The DYSAN framework in chapter IV.1 aims at overcoming this limitation by maintaining the transformed signal as close as possible to the raw signal.

Dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks. The chapter IV.1 proposed DySAN, a framework that sanitizes motion sensor data against unwanted sensitive inferences while limiting the loss of accuracy on the physical activity monitoring. Our approach is inspired from the framework of GANs to sanitize the sensor data for the purpose of ensuring a good trade-off between utility and privacy. More precisely, by learning in a competitive manner several networks, DySAN is able to build models that sanitize motion data against inferences on a specified sensitive attribute (e.g., gender) while maintaining an accurate activity recognition. DySAN builds various sanitizing models, characterized by different sets of hyperparameters in the global loss function, to propose a transfer learning scheme over time by dynamically selecting the model which provides the best utility and privacy trade-off according to the incoming data. Although we have shown that it is possible to run this solution on smartphones, a calibration phase is needed to label the user data. Experiments conducted on real datasets demonstrated that DySAN can drastically limit the gender inference up to 41% (from 98% with raw data to 57% with sanitized data) while only reducing the accuracy of activity recognition by 3% (from 95% with raw data to 92% with sanitized data).

DySAN framework is constrained to sanitize only one feature at one time. We investigated the possibility of extending DySAN to take into account multiple sensitive attributes. Our preliminary results by adding two discriminators accounted for in the loss function of the sanitizer's training are encouraging. However, we are limited by the small size of the available datasets. Indeed, making the sanitizing models more complex requires more data to capture the specificity of each use case. Moreover, having numerous sensitive attributes to sanitize suppose either a neural network assigned to each sensitive attribute or a Multi-task neural network [67]. Optimizing simultaneously several objective functions can be tough when the different objectives are conflicting. In this case, the Pareto solutions (set of optimal solutions) to a multi-objective problem are those for which the performance of one objective can only be improved by deteriorating the performance of another objective [67]. Another limitation is the dependency of the features between each other. If one of the sensitive attributes is highly correlated to the activity, finding the optimal solution is all the more difficult. Considering the limitations on multiple sensitive attributes protection, collaborative learning is a way to overcome any sensitive inference on the data at the cloud level.

Privacy Assessment of FL using Private Personalized Layers. The chapter IV.2 is a preliminary work on FL, where we quantified the utility and privacy trade-off of a FL scheme using private personalized layers. While this scheme has been proposed as local adaptation to improve the accuracy of the model through local personalization with non-identical client distributions, it has also the advantage to minimize the information on the model exchanged with the server. However, the privacy of such a scheme has never been quantified. Our evaluations using motion sensor dataset tended to show that personalized layers speed up the convergence of the model and slightly improve the accuracy for all users compared to a standard FL scheme while better preventing both attribute and membership inferences compared to a FL scheme using local differential privacy.

This numerical evaluation needs further exploration to extend this contribution by using several datasets and several ML models. Another way is to use a hybrid approach such as combining personalizing layers sharing with LDP methods. It is known that preserving utility with LDP methods is challenging [161] due to the fact that the magnitude of the random noise introduced is often comparable with the magnitude of the signal in the data. Combining LDP and personalizing layers could lower the quantity of noise introduced while keeping privacy guarantees.

V.2 Discussion and research openings

In the introduction of this work we wondered if it would be possible to design a ML framework able to protect motion sensor data from sensitive inference information while ensuring utility of this data for healthcare applications. The thesis provides a positive answer of this question but under certain conditions on the privacy-utility trade-off.

Indeed, each contribution provided hypotheses on the privacy threats that each framework designed intends to address. Future research direction should tend to systems that address several privacy issues in the same time, whether at the user's level on the smartphone, during communications or at the server's level, whether for data or model leakages, and so on. Achieving all this desired privacy properties probably requires composing many different strategies into an end-to-end system. A recent publication proposes PriMIA [162] an open-source framework that uses differentially private federated model training with encrypted aggregation of model updates as well as encrypted remote inference applied on medical imaging analysis. This is one of the first approaches that proposes an end-to-end framework that can potentially be deployed in practice.

Moreover, there is a need for a more thorough assessment on how defences operate in practice, facing realistic use cases and datasets rather than the standard public ones, with clinical data that record gait users with pathologies. This thesis is also indeed limited by this aspect, as we could only assess our contributions on benchmark datasets. Activity detection is an important aspect of gait monitoring that has been explored, but other type of monitoring such as event detection (*i.e.*, stance, fall) should be considered with the use of Recurrent Neural Networks as they have demonstrated a great accuracy in some problems that require analyzing sequential inputs [167].

The concept of privacy-utility trade-off also needs to be further explored and specifically when this trade-off is considered as viable. Indeed specifically optimizing this trade-off may impacts other aspects such as fairness, robustness and efficiency. In this thesis, we specifically focused on the personalization of the privacy framework, by observing that maximising a utility metric in average for an entire group of users, could lead to high inequalities between users. In practice, a same framework could provide a strong privacy protection to one user and a lower privacy protection to another one. Chapter IV specifically focuses on this personalizing aspect by attempting to provide privacy protection to every user. To go further in this direction, it could be interesting to design flexible solutions that allow a certain trade-off involving utility, efficiency, privacy, and ability to address different constraints and requirements following the application considered. In practice, to provide sufficient privacy guarantees, the privacy framework must understand the user's privacy needs in relation to the specific analysis task and data collection procedure. The framework could be modified to allow each user to specify what inferences are allowed or not. These restrictions could be processed on device, by sharing with the server only the information allowed by the user. Future works should then develop methods to incorporate user preferences into the privacy framework, and adapt it to potential updates made by a user on his own privacy policy. Collaborative learning methods seem particularly adapted to this aspect of personalizing privacy, because it allows the preservation of data sovereignty and application of local governance for each user. Exploring this aspect could bring privacy ML framework closer to the legal requirements of the GDPR regarding the consent of each individual (see Art.4(11) [1]), which must be *freely given*, *specific*, *informed* and *unambiguous*.

Chapter VI

Appendices

A Extraction from databases in state-of-the-art Section II.2

Database	Search string	Records
ACM	[[Abstract: gait] OR [Abstract: actimetry] OR [Abstract: actigraphy] OR [Abstract: walk]] AND [[[Abstract: smartphone] OR [Abstract: wearable] OR [Abstract: iot]] AND [[Abstract: "chronic disease"] OR [Abstract: rehabilitation] OR [Abstract: medicine]] AND [[Abstract: validity] OR [Abstract: reliability] OR [Abstract: reproductibility or validation] OR [Publication Title: gait] OR [Publication Title: actimetry] OR [Publication Title: actigraphy] OR [Publication Title: walk]] AND [[Publication Title: smartphone] OR [Publication Title: wearable] OR [Publication Title: iot] AND [Publication Title: "chronic disease"] OR [Publication Title: rehabilitation] OR [Publication Title: medicine]] AND [[Publication Title: validity] OR [Publication Title: reliability] OR [Publication Title: reproductibility or validation]] AND [Publication Date: (01/01/2010 TO 10/31/2020)]	17
Cochrane	((gait OR actimetry OR actigraphy OR walk) AND (smartphone OR wearable OR iot) AND ("chronic disease" OR rehabilitation OR medicine) AND (validity OR reliability OR reproductibility OR validation)) in Title Abstract Keyword - between Jan 2010 and October 2020	15
DBLB	(gait | walk | actimetry) (smartphone | device | iot) (valid | rehabilitation)	31
IEEE Xplore	((gait OR actimetry OR actigraphy OR walk) AND (smartphone OR wearable OR iot) AND ("chronic disease" OR rehabilitation OR medicine) AND (validity OR reliability OR reproductibility or validation))	54
PubMed	((gait OR actimetry OR actigraphy OR walk) AND (smartphone OR wearable OR iot) AND ("chronic disease" OR rehabilitation OR medicine) AND (validity OR reliability OR reproductibility or validation)) Filters: from 2010 - 2020	52
Scholar	title:(gait smartphone "wearable device" rehabilitation validity)	1010
ScienceDirect #1	((gait OR actimetry) AND (smartphone OR iot) AND ("chronic disease" OR medicine) AND (validity OR validation))	3
ScienceDirect #2	((gait OR walk) AND (smartphone OR wearable) AND (rehabilitation OR medicine) AND (validity OR reliability))	10
ScienceDirect #3	((gait OR walk) AND (smartphone OR iot) AND AND ("chronic disease" OR medicine) AND (validity OR validation))	1
ScienceDirect #4	((gait OR walk) AND (smartphone OR wearable) AND AND (rehabilitation OR medicine) AND (validity OR validation))	16
ScienceDirect #5	((gait OR actimetry OR walk) AND (smartphone OR wearable OR iot) AND rehabilitation AND validation)	12
SCOPUS	TITLE-ABS-KEY(((gait OR actimetry OR actigraphy OR walk) AND (smartphone OR wearable OR iot) AND ("chronic disease" OR rehabilitation OR medicine) AND (validity OR reliability OR reproductibility OR validation))) AND PUBYEAR $\geq$ 2010 AND PUBYEAR $\leq$ 2020	155
Web of Science	(TS = ((gait OR actimetry OR actigraphy OR walk) AND (smartphone OR wearable OR iot) AND ("chronic disease" OR rehabilitation OR medicine) AND (validity OR reliability OR reproductibility OR validation))) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article) Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=2010-2020	148

TABLE 1 – Search term strategy.

B Criteria selection for state-of-the-art Section II.2

Author	Year	Pathology	Cohort size	Length data collection	Condition data collection
Salarian et al. [269]	2010	Parkinson	12	min	Laboratory
Dobkin et al. [76]	2011	Stroke	12	min (Lab), days (FL)	Both
Kozey-Keadle et al. [176]	2011	Obesity	20	hours	Free living
Munguía-Izquierdo et al. [218]	2012	Fibromyalgia	25	min	Laboratory
Item-Glatthorn et al. [148]	2012	Osteoarthritis	26	min	Laboratory
Grimpampi et al. [125]	2013	Hemiplegia/Parkinson	24	min	Laboratory
Schwenk et al. [280]	2014	Dementia	77	days	Free living
Juen et al. [158]	2014	Lung disease	30	min	Laboratory
Juen et al. [159]	2014	Lung disease	25	min	Laboratory
Sprint et al. [291]	2015	Diverse	20	min	Laboratory
Capela et al. [52]	2015	Lung disease	15	min	laboratory
Schwenk et al. [279]	2016	Cancer	22	hours	laboratory
Isho et al. [146]	2015	Stroke	24	min	Laboratory
Wuest et al. [335]	2016	Stroke	26	min	Laboratory
Raknim et al. [252]	2016	Parkinson	1	years	Free living
Ferrari et al. [96]	2016	Parkinson	14	min	Laboratory
Brinkløv et al. [48]	2016	Diabete	27	min	Laboratory
El-Gohary et al. [85]	2017	Multiple sclerosis	52	min	Laboratory
Ilias et al. [145]	2017	Parkinson	19	min	Laboratory
Maqbool et al. [202]	2017	Amputee	2	min	Laboratory
Terrier et al. [302]	2017	Chronic Pain	66	weeks	Both
Rogan et al. [259]	2017	Old-age	23	min	Laboratory
Chiu et al. [62]	2017	Ankle instability	15	min	Laboratory
Cheng et al. [59]	2017	Cardiopulmonary disease	25	min	Laboratory
Kobsar et al. [173]	2017	Osteoarthritis	39	months	Laboratory
McGinnis et al. [205]	2017	Multiple sclerosis	30	min	Laboratory
Lipsmeier et al. [191]	2018	Parkinson	44	months	Free living
Kleiner et al. [171]	2018	Parkinson	30	min	Laboratory
Carpinella et al. [54]	2018	Diverse	30	min	Laboratory
Jayaraman et al. [154]	2018	Spinal Cord Injury	18	hours	Laboratory
Jang et al. [153]	2018	Old-age	22	years	Free living
Derungs et al. [74]	2018	Hemiparesis	11	weeks	Free living
Mileti et al. [212]	2018	Parkinson	26	min	Laboratory
Aich et al. [9]	2018	Parkinson	51	min	Laboratory
Cheong et al. [60]	2018	Cancer	102	months	Free living
Ata et al. [29]	2018	Artery disease	114	min	Laboratory
Kim et al. [168]	2018	Parkinson	32	min	Laboratory
Vadnerkar et al. [315]	2018	Old-age	16	min	Laboratory
Rosario et al. [73]	2018	Cardiac disease	66	months	Free living
Lemoyne et al. [185]	2018	Hemiplegia	1	min	Laboratory
Dasmahapatra et al. [71]	2018	Multiple Sclerosis	114	weeks	Free living
Schliessmann et al. [277]	2018	Diverse	41	min	Laboratory
Ummels et al. [314]	2018	Diverse	130	years	Laboratory
Banky et al. [33]	2019	Diverse	35	hours	Laboratory
Flachenecker et al. [99]	2019	Multiple sclerosis	102	min	Laboratory
Gadaleta et al. [108]	2019	Parkinson	71	min	Laboratory
Teufl et al. [303]	2019	Arthroplasty	20	min	Laboratory
Angelini et al. [18]	2019	Multiple sclerosis	26	min	Laboratory
Antos et al. [21]	2019	Old-age	20	min	Laboratory
Compagnat et al. [64]	2019	Stroke	35	min	Laboratory
Newman et al. [227]	2020	Brain injury	12	min	Laboratory
Ullrich et al. [313]	2020	Parkinson	128	min	Both
Wang et al. [322]	2020	Post Sternotomy	22	min	Laboratory
Pavon et al. [243]	2020	Disability	46	days	Laboratory
Arcuria et al. [24]	2020	Cerebellar ataxia	40	min	Laboratory
Erb et al. [87]	2020	Parkinson	34	weeks	Free Living
Aich et al. [8]	2020	Parkinson	48	min	Laboratory
Rubin et al. [263]	2020	Diverse	78	min	Laboratory
Henriksen et al. [133]	2020	Obesity	16	years	Free living
Shema-Shiratzky et al. [283]	2020	Multiple Sclerosis	44	min	Both
Abdollahi et al. [5]	2020	Chronic pain	94	min	Laboratory
Kim et al. [169]	2020	Amputation	17	min	Laboratory
Lemay et al. [184]	2020	Spinal cord injury	18	min	Laboratory
Meisel et al. [208]	2020	Epilepsy	69	months	Laboratory
Fantozzi et al. [91]	2020	Old-age	9	min	Laboratory
Zhai et al. [346]	2020	Multiple Sclerosis	67	min (Lab), weeks (FL)	Both
Revi et al. [256]	2020	Stroke	5	min	Laboratory
Compagnat et al. [63]	2020	Stroke	26	min	Laboratory
Furtado et al. [106]	2020	Amputation	34	hours (Lab), weeks (FL)	Both
Na et al. [220]	2020	Osteoarthritis	39	min	Laboratory

TABLE 2 – Data acquisition criteria through the 70 selected papers. Abbreviations used in column "Length of data collection" : min (t<1 hour), hours (1≤t<24 hours), days (1≤t<7 days), weeks (1≤t<4 weeks), months (1≤t<12 months), year (t≥1 year). Finally, the cohort size is given in number of patients.

Author	No. of device(s)	Sensor type(s)	Location of device(s)	Sensor model, Brand
Salarian et al. [269]	7 (IMU)	A,G	Forearms, shanks, thighs, sternum	Physilogs, BioAGM
Dobkin et al. [76]	2 (S)	A	Both ankles	GCDC, LLC
Kozey-Keadle et al. [176]	2 (S)	A	Right leg, right side of the hip	activPAL, PALF GT3X, ActiGraph
Munguía-Izquierdo et al. [218]	1 (IMU)	A,O	Arm	SenseWear, Bodymedia
Item-Glatthorn et al. [148]	5 (S)	A	Chest, thigh, forefoot	MiniSun, IDEEA
Grimpampi et al. [125]	1 (IMU)	A,G	Lumbar spine	Freesense, Sensorize
Schwenk et al. [280]	1 (IMU)	A,G	chest	Physilog, GaitUp
Juen et al. [158]	1 (SPHN)	A	pants pocket or in fanny pack	Galaxy Ace, Samsung
Juen et al. [159]	2 (SPHN and S)	A	L3 vertebra	Galaxy Ace/4, Samsung
Sprint et al. [291]	3 (IMU)	A,G	Lumbar spine, shank	Shimmer3, Shimmer
Capela et al. [52]	1 (SPHN)	A,G,M	Rear pocket	Z10, BlackBerry
Schwenk et al. [279]	5 (IMU)	A,G,M	Shank, thigh, lower back	LegSys, BioSensic
Isho et al. [146]	1 (SPHN)	A	Torso	Xperia Ray SO-03C, Sony
Wuest et al. [335]	8 (IMU)	A,G	Wrists, shanks, trunk, feet, back	Physilog, GaitUp
Raknim et al. [252]	1 (SPHN)	A	Free (pocket, during phone call, on the bag during walk)	HTC and Samsung
Ferrari et al. [96]	2 (IMU)	A,G	Shoes	EXLs1 and EXLs3, EXEL
Brinkløv et al. [48]	1 (SPHN)	A	Pants pocket, jacket pocket	Iphone 5C, Apple
El-Gohary et al. [85]	3 (IMU)	A,G	Lumbar vertebra, feet, ankles	Opal, APDM
Ilias et al. [145]	4 (IMU)	A,G	Upper, lower limbs, wrists, legs	Shimmer3, Shimmer
Maqbool et al. [202]	1 (IMU)	A,G	Shank	MPU 6050, InvenSense
Terrier et al. [302]	1 (S)	A	Right hip	wGT3X-BT, ActiGraph
Rogan et al. [259]	1 (IMU)	A,G	Lateral malleolus	RehaWatch, Hasomed
Chiu et al. [62]	1 (SPHN)	A	Shin	Zenfone 2, ASUS
Cheng et al. [59]	1 (SPHN)	A	Carried in 'fanny pack	Galaxy S5, Samsung Optimus Zone2, LG
Kobsar et al. [173]	4 (IMU)	A,G	Foot, shank, thigh, lower back	iNEMO, STmicroelectronics
McGinnis et al. [205]	5 (IMU)	A	Sacrum, thighs, shanks	BioStampRC, MC10
Lipsmeier et al. [191]	1 (SPHN)	A,G,M,O	Hand, trouser pocket, belt	Galaxy S3 mini, Samsung
Kleiner et al. [171]	1 (IMU)	A,G,M	L5 verterbra	BTS G-walk, BTS G-Sensor
Carpinella et al. [54]	1 (IMU)	A,G,M	Sternum	MTw, Xsens
Jayaraman et al. [154]	4 (S)	A,O	Arm, waist, ankle	wGT3X-BT, ActiGraph Metria-IH1, Vandrico
Jang et al. [153]	1 (IMU)	A,O	Wrist	Mi band 2, Xiaomi
Derungs et al. [74]	6 (IMU)	A,G,M	Wrists, arms, thighs	Shimmer3, Shimmer
Mileti et al. [212]	10 (IMU and S)	A,G,M,O	feet	Mtw, MTw, Xsens
Aich et al. [9]	2 (S)	A	knees	Fit Meter, Fit.Life
Cheong et al. [60]	1 (IMU)	A	Wrists	Urban S, Partron Co
Ata et al. [29]	2 (SPHN and S)	A	Hand, hip	iPhones SE/6/7/7+, Apple GT9X, ActiGraph
Kim et al. [168]	3 (SPHN)	A,G	Waist, pocket, ankle	Nexus 5, Google
Vadnerkar et al. [315]	1 (IMU)	A,G	Feet	Shimmer 2r, Shimmer
Rosario et al. [73]	1 (SPHN)	A,G	Trouser pocket	Galaxy S3, Samsung
Lemoyne et al. [185]	1 (SPHN)	A	Malleolus	iPhone, Apple
Dasmahapatra et al. [71]	1 (S)	A	Belt, pocket or bra	Fitbit One, Fitbit
Schliessmann et al. [277]	2 (IMU)	A,G,M	Feet	RehaGait, HASOMED GmbH UP24, Jawbone
Ummels et al. [314]	9 (IMU and S)	other	Leg, belt, wrist	Lumoback, Lumo Bodytech Moves, ProtoGeo Oy Accupedo, Corusen LLC Walking Style X, Omron
Banky et al. [33]	1 (SPHN)	G		Galaxy S5, Samsung
Flachenecker et al. [99]	2 (IMU)	A,G	Shoes	Shimmer 3, Shimmer
Gadaleta et al. [108]	3 (IMU)	A,G,M	L5 lumbar vertebrae, ankles	Opal, APDM
Teufl et al. [303]	7 (IMU)	A,G	Pelvis, both foot, both thighs	MTw Awinda, Xsens
Angelini et al. [18]	3 (IMU)	A,G	L5 lumbar vertebra, ankles	MTw Xsens Opal, APDM
Antos et al. [21]	2 (S and SPHN)	A,G	Waist, wrist	Nexus 5 , Google wGT3X-BT, Actigraph
Compagnat et al. [64]	9 (S)	A,O	Wrists, ankles, hip, arm, neck	GT3x, Actigraph Sensewear, Body Media
Newman et al. [227]	1 (IMU)	A,G	Interclavicular notch	Opal, APDM
Ullrich et al. [313]	3 IMU	A,G	Ankles, shoes	Shimmer2R, Shimmer
Wang et al. [322]	2 (IMU)	A,G	Pectoralis major	BioStampRC, MC10
Pavon et al. [243]	2 (S)	A	Ankle	GT3x+, ActiGraph
Arcuria et al. [24]	1 (SPHN)	A	Breastbone	Galaxy J3, Samsung
Erb et al. [87]	7 to 16 (IMU)	A,G,M,O	wrists, torso, thigh, feet	Shimmer, Shimmer
Aich et al. [8]	2 (S)	A	Knees	Fit Meter, Fit. Life
Rubin et al. [263]	1 (SPHN)	A,G	Pants pocket, belt	iPhone 6, Apple
Henriksen et al. [133]	1 (IMU)	A,O	Wrist	M430 AT, Polar
Shema-Shiratzky et al. [283]	1 (IMU)	A	Lower Back	Opal, APDM and AX3, Axivity
Abdollahi et al. [5]	1 (IMU)	A,G	Sternum	9DOF Razor IMU, Sparkfun
Kim et al. [169]	2 (IMU)	A,G	Shoe, ankle	GT9X Link, ActiGraph
Lemay et al. [184]	5 (IMU)	A,G,O	Feet, shanks, sacrum	Physilog, GaitUp
Meisel et al. [208]	1 (S)	A,O	Wrist or ankle	E4, Empatica
Fantozzi et al. [91]	5 (IMU)	A,G,M	Trunk, pelvis, thigh, shank, foot	Opal, APDM
Zhai et al. [346]	2 (SPHN and S)	A	Wrist, pocket	Galaxy S4 mini, Samsung GT3X+, ActiGraph
Revi et al. [256]	3 (IMU)	A	Shank, thigh, pelvis	MTw Awinda, Xsens
Compagnat et al. [63]	2 (S)	A	Non-paretic hip	GT3x, ActiGraph
Furtado et al. [106]	1 (S)	A	L5 lumbar vertebrae within the pocket of a belt	AX3, Axivity
Na et al. [220]	5 (IMU)	A,G	Femur, tibia, pelvis, sacral ridge	3D Myomotion, Noraxon

TABLE 3 – Criteria related to commercial wearable devices through the 70 selected papers. Abbreviations used in column "No. of device(s)": IMU (Inertial Motion Unit), S (Sensor), SPHN (Smartphone). Abbreviations used in column "Sensor Type(s)" : A (accelerometer), G (gyroscope), M (magnetometer), O (others).

Author	Ground truth method	Gait descriptors	# of descriptors	Evaluation method	Evaluation outcomes
Salarian et al. [269]	controls, medical	high	20	stats + test	p-value<0.023
Dobkin et al. [76]	controls, metrologic	medium	8	ML + test	r=0.98
Kozey-Keadle et al. [176]	expert	high	3	stats	R^2=0.94
Munguía-Izquierdo et al. [218]	med device	high	1	stats + test	r=0.87-0.99
Item-Glatthorn et al. [148]	metrologic	high	6	stats + test	ICC =0.815-0.997
Grimpampi et al. [125]	metrologic	low, medium	3	stats + test	r=0.74-0.87
Schwenk et al. [280]	controls, user	high	9	stats + test	AUC=0.77, sen/spe=72%/76%
Juen et al. [158]	medical	medium	8	ML	acc=89.22-94.13%
Juen et al. [159]	med device	medium	9	ML	error<10.2%
Sprint et al. [291]	medical	medium,high	18	ML + test	r=0.97
Capela et al. [52]	expert	high	10	stats	time difference=0.014 s
Schwenk et al. [279]	controls, user	high	6	LM + test	p-value<0.022
Isho et al. [146]	controls, user	medium	3	ML + test	AUC=0.745
Wuest et al. [335]	controls, medical	high	13	stats + test	p-value<0.02
Raknim et al. [252]	controls	high	2	ML	acc=94%
Ferrari et al. [96]	metrologic	high	4	LM + test	error=2,9%
Brinkløv et al. [48]	med device	medium	6	LM + test	R^2=0.45-0.60
El-Gohary et al. [85]	metrologic, controls	high	7	stats + test	r=0.592-0.992
Ilias et al. [145]	expert	medium	152	ML + test	r=0.78-0.79
Maqbool et al. [202]	metrologic, controls	high	1	stats	time difference=50 ms
Terrier et al. [302]	controls, medical	high	4	LM + stats	R^2=0.44
Rogan et al. [259]	metrologic	high	6	stats + test	p-value<0.05
Chiu et al. [62]	controls	medium	1	stats + test	p-value<0.027
Cheng et al. [59]	med device, medical	medium,high	10	ML	NA
Kobsar et al. [173]	medical	medium	38	LM + test	acc=74-81.7%
McGinnis et al. [205]	metrologic, controls	medium	32	ML + test	speed difference=0.12-0.16 m/s
Lipsmeier et al. [191]	controls, medical	high	6	ML + test	p-value<0.055
Kleiner et al. [171]	metrologic, medical	high	1	stats	time difference=0.585 s
Carpinella et al. [54]	medical, controls	high	5	stats + test	r=-0.367-0.536
Jayaraman et al. [154]	expert, metrologic	high	3	stats + test	p-value<0.05
Jang et al. [153]	controls	high	5	stats + test	p-value<0.02
Derungs et al. [74]	expert	medium	8	LM + test	sen/spe=80%/94%
Mileti et al. [212]	controls, medical	low	3	ML + test	AUC=0,48-0,98
Aich et al. [9]	metrologic, controls	high	28	ML	acc=88%
Cheong et al. [60]	controls	high	1	stats + test	p-value<0.04
Ata et al. [29]	expert, med device	high	3	stats	R^2=0.9-0.92
Kim et al. [168]	expert	medium	8	ML	sen/spe=93.8%/90.1%
Vadnerkar et al. [315]	expert	low	1	LM + test	acc=84%, sen/spe=75.9%/95.9%
Rosario et al. [73]	controls, medical	high	2	stats + test	r=0.472
Lemoyne et al. [185]	controls	high	5	stats + test	p-value<0.05
Dasmahapatra et al. [71]	controls, medical	high	6	LM + test	p-value<0.05
Schliessmann et al. [277]	controls	high	4	stats + test	p-value<0.05
Ummels et al. [314]	metrologic	high	1	stats + test	r=-0.02-0.33
Banky et al. [33]	metrologic, controls	low	3	stats + test	r=0.8
Flachenecker et al. [99]	controls, medical	high	8	stats + test	r=-0,583-0,668
Gadaleta et al. [108]	metrologic	low	24	ML	bias=-0.012-0.000, IQR=0.004-0.032
Teufl et al. [303]	metrologic, controls	high	10	ML + test	acc=0.87-0.97
Angelini et al. [18]	expert, controls	high	14	stats + test	p-value<0.05
Antos et al. [21]	expert, controls	medium	56	ML + test	acc=0.90-0.95
Compagnat et al. [64]	expert	high	2	stats + test	p-value<0.05
Newman et al. [227]	controls, medical	high	9	stats + test	p-value<0.05
Ullrich et al. [313]	expert	medium	7	stats + test	sen/spe=98%/96%
Wang et al. [322]	controls	medium	1	stats + test	p-value<0.05
Pavon et al. [243]	controls, medical	high	3	stats + test	p-value<0.16
Arcuria et al. [24]	metrologic, controls, medical	high	1	stats + test	r=-0.72-0.91
Erb et al. [87]	user, expert	high	2	stats + test	FN=35%, FP=15%
Aich et al. [8]	metrologic, controls, medical	high	5	ML	acc=88.46%
Rubin et al. [263]	med device	high	1	stats + test	R^2=0.72
Henriksen et al. [133]	med device	high	4	stats	r=0.446-0.925
Shema-Shiratzky et al. [283]	controls, expert	high	5	stats + test	p-value<0.05
Abdollahi et al. [5]	medical	medium	920	ML	acc=60-75%
Kim et al. [169]	controls	high	5	stats + test	p<0.05
Lemay et al. [184]	medical, controls	high	6	LM + test	r=-0.49-0.498
Meisel et al. [208]	expert	low	6	ML + test	acc=43%
Fantozzi et al. [91]	controls	high	14	LM + test	NA
Zhai et al. [346]	med device, controls, medical	medium	14	stats + test	r=0.43-0.605
Revi et al. [256]	metrologic	high	8	stats	R^2=0.90-0.93
Compagnat et al. [63]	med device	high	1	stats + test	r=0.44-0.87
Furtado et al. [106]	metrologic, controls, medical	medium,high	10	stats + test	p-value<0.024
Na et al. [220]	metrologic, controls	high	6	stats + test	p-value<0.04

TABLE 4 – Evaluation criteria through the 70 selected papers. Abbreviations used in column "Evaluation method" : stats (descriptive statistics), stats + test (descriptive statistics + statistical tests), LM + test (linear models + statistical tests), ML (machine learning), ML+test (machine learning + statistical tests). Abbreviations used in column "Evaluation outcomes" : r (correlation coefficient), R^2 (coefficient of determination), ICC (intraclass correlation coefficient), AUC (area under curve, sen (sensitivity), spe (specificity), IQR (interquartile range), FN (false negatives), FP (false positives), acc (accuracy).

Author	Task	Model type	Training size	# of descriptors	Outcome
Dobkin et al. [76]	Speed prediction	Naive Bayes	NA	24	r=0.98
Juen et al. [158]	Healthy/Patient	SVM	10-20	8	accuracy=89.22-94.13%
Juen et al. [159]	Speed prediction Distance prediction	GPR NN SVM	24	60	error rate = 2.51% error rate = 10.2%
Sprint et al. [291]	FIM motor score prediction	SVM RF	19	18	NRMSE = 10%-30%
Raknim et al. [252]	Step length estimation Before/After PD	SVM	1	2	accuracy=98% accuracy=94%
Ilias et al. [145]	Motor function prediction	SVM	6	152	RMSE = 0.46-0.70 r=0.78-0.79
Cheng et al. [59]	3 pulmonary severity stages	SVM	22-25	10	NA
McGinnis et al. [205]	Walking speed	SVM	16	32	RMSE = 10%-20%
Lipsmeier et al. [191]	Activities	LSTM	44	6 (*n)	accuracy=98%
Mileti et al. [212]	4 Gait phases	HMM	1-11	3 (*n)	AUC=0.48-0.98 sens= 80%-100 spe=70%-90% goodness Index = 10%-40%
Aich et al. [9]	Healthy/Patient	SVM Decision tree Naive Bayes kNN	36	28	accuracy=91.42% sens/spe = 90.9%/91.2%
Kim et al. [168]	Walking/Freezing	CNN	29	8 (*n)	f1-score = 91.8 sen/spe=93.8%/90.1%
Vadnerkar et al. [315]	Gait quality	ROC decision boundary	8	1	accuracy=84% sen/spe=75.9%/95.9%
Gadaleta et al. [108]	Right/Left foot events	CNN	138	24 (*n)	bias=-0.012-0.000 IQR=0.004-0.032
Teufl et al. [303]	Healthy/Patient	SVM RF	40	10	accuracy=87-97%
Antos et al. [21]	With/without assistance	SVM Naive Bayes Logistic regression LDA kNN	1-13	56	accuracy=90-95%
Aich et al. [8]	Healthy/Patient	SVM Naive Bayes Decision tree	62	10	accuracy=88.5% sens/spe=92.9%/90.9%
Abdollahi et al. [5]	Risk of disability	SVM Perceptron	93	920	accuracy=60-75%
Meisel et al. [208]	Seizure/Healthy	LSTM	68	6 (*n)	accuracy=43%

TABLE 5 – Selection of papers that use machine learning methods in validation. Abbreviations used in column "Model type": SVM (support vector machine), GPR (gaussian process regression), NN (neural network), RF (random forest), LSTM (long short time memory), HMM (hidden markov model), kNN (k-nearest neighbors), CNN (convolutional neural network), ROC (receiver operating characteristic), LDA (linear discriminant analysis). Abbreviations used in column "Outcome": r (correlation coefficient), NRMSE (normalized root mean square error), RMSE (root mean square error), AUC (area under curve), sens (sensitivity), spe (specificity), IQR (interquartile range). Studies that use raw data as input have a number of descriptors that corresponds to the number of sensors and/or axes multiplied by the length of the recorded data (n).

Bibliography

[1] *2018 reform of EU data protection rules*. European Commission. May 25, 2018. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj (visited on 07/17/2021).

[2] *2020 Q1 Report Data Breach QuickView*. Risk Based Security. May 11, 2020. URL: https://library.cyentia.com/report/report_003567.html (visited on 07/20/2021).

[3] A. Abadleh, E. Al-Hawari, E. Alkafaween, and H. Al-Sawalqah. "Step detection algorithm for accurate distance estimation using dynamic step length". In: *MDM*. 2017, pp. 324–327.

[4] S. Abbate, M. Avvenuti, F. Bonatesta, G. Cola, P. Corsini, and A. Vecchio. "A smartphone-based fall detection system". In: *Pervasive and Mobile Computing* 8.6 (2012). Special Issue on Pervasive Healthcare, pp. 883–899.

[5] M. Abdollahi, S. Ashouri, M. Abedi, N. Azadeh-Fard, M. Parnianpour, K. Khalaf, and E. Rashedi. "Using a Motion Sensor to Categorize Nonspecific Low Back Pain Patients: A Machine Learning Approach". In: *Sensors* 20.12 (2020), p. 3600.

[6] G. Acs and C. Castelluccia. "A case study: Privacy preserving release of spatio-temporal density in paris". In: *KDD*. 2014, pp. 1679–1688.

[7] B. Aguiar, T. Rocha, J. Silva, and I. Sousa. "Accelerometer-based fall detection for smartphones". In: *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2014, pp. 1–6.

[8] S. Aich, P. M. Pradhan, S. Chakraborty, H.-C. Kim, H.-T. Kim, H.-G. Lee, I. H. Kim, M.-i. Joo, S. Jong Seong, and J. Park. "Design of a Machine Learning-Assisted Wearable Accelerometer-Based Automated System for Studying the Effect of Dopaminergic Medicine on Gait Characteristics of Parkinson's Patients". In: *Journal of healthcare engineering* 2020 (2020).

[9] S. Aich, P. M. Pradhan, J. Park, N. Sethi, V. S. S. Vathsa, and H.-C. Kim. "A validation study of freezing of gait (FoG) detection and machine-learning-based FoG prediction using estimated gait characteristics with a wearable accelerometer". In: *Sensors* 18.10 (2018), p. 3287.

[10] U. Aïvodji, F. Bidet, S. Gambs, R. C. Ngueveu, and A. Tapp. "Agnostic data debiasing through a local sanitizer learnt from an adversarial network approach". In: *arXiv preprint arXiv:1906.07858* (2019).

[11] Y. Al-Issa, M. Ottom, and A. Tamrawi. "eHealth Cloud Security Challenges: A Survey". In: *Journal of Healthcare Engineering* 2019 (2019).

[12] A. Al-Mahmood and M. O. Agyeman. "On wearable devices for motivating patients with upper limb disability via gaming and home rehabilitation". In: *2018 Third International Conference on Fog and Mobile Edge Computing (FMEC)*. 2018, pp. 155–162.

[13] M. Alaqtash, T. Sarkodie-Gyan, H. Yu, O. Fuentes, R. Brower, and A. Abdelgawad. "Automatic classification of pathological gait patterns using ground reaction forces and machine learning algorithms". In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2011 (Aug. 2011), pp. 453–7.

[14] A. S. Alharthi, S. U. Yunas, and K. Ozanyan. "Deep Learning for Monitoring of Human Gait: A Review". In: *IEEE Sensors Journal* 19 (2019), pp. 9575–9591.

[15] R. Altilio, A. Rossetti, Q. Fang, X. Gu, and M. Panella. "A comparison of machine learning classifiers for smartphone-based gait analysis". In: *Medical & Biological Engineering & Computing* 59.3 (2021), pp. 535–546.

[16] M. Alzantot, S. Chakraborty, and M. Srivastava. "Sensegen: A deep learning architecture for synthetic sensor data generation". In: *PerCom Workshops*. 2017, pp. 188–193.

[17] *Amazon Elastic Compute Cloud (Amazon EC2)*. http://aws.amazon.com/ec2..

[18] L. Angelini, I. Carpinella, D. Cattaneo, M. Ferrarin, E. Gervasoni, B. Sharrack, D. Paling, K. P. S. Nair, and C. Mazzà. "Is a wearable sensor-based characterisation of gait robust enough to overcome differences between measurement protocols? A multicentric pragmatic study in patients with multiple sclerosis". In: *Sensors* 20.1 (2020), p. 79.

[19] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. "A Public Domain Dataset for Human Activity Recognition using Smartphones." In: *ESANN*. 2013.

[20] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz. "Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine". In: *Ambient Assisted Living and Home Care*. Ed. by J. Bravo, R. Hervás, and M. Rodríguez. Springer Berlin Heidelberg, 2012, pp. 216–223.

[21] S. A. Antos, M. K. Danilovich, A. R. Eisenstein, K. E. Gordon, and K. P. Kording. "Smartwatches can detect walker and cane use in older adults". In: *Innovation in aging* 3.1 (2019), igz008.

[22] N. J. Apthorpe, D. Reisman, and N. Feamster. "A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic". In: *CoRR* (2017).

[23] D. Aranki and R. Bajcsy. "Private Disclosure of Information in Health Tele-monitoring". In: *CoRR* abs/1504.07313 (2015).

[24] G. Arcuria, C. Marcotulli, R. Amuso, G. Dattilo, C. Galasso, F. Pierelli, and C. Casali. "Developing a smartphone application, triaxial accelerometer-based, to quantify static and dynamic balance deficits in patients with cerebellar ataxias". In: *Journal of neurology* 267.3 (2020), pp. 625–639.

[25] M. G. Arivazhagan, V. Aggarwal, A. K. Singh, and S. Choudhary. "Federated Learning with Personalization Layers". In: *CoRR* (2019).

[26] A. Artese, D. Ehley, A. Sutin, and A. Terracciano. "Personality and actigraphy-measured physical activity in older adults". In: *Psychology and Aging* 32 (Mar. 2017), pp. 131–138.

[27] L. J. M. Aslett, P. M. Esperança, and C. C. Holmes. *Encrypted statistical machine learning: new privacy preserving methods*. 2015.

[28] S. Assous and B. Boashash. "Evaluation of the modified S-transform for time-frequency synchrony analysis and source localisation". In: *EURASIP Journal on Applied Signal Processing* 2012.1 (Dec. 2012).

[29] R. Ata, N. Gandhi, H. Rasmussen, O. El-Gabalawy, S. Gutierrez, A. Ahmad, S. Suresh, R. Ravi, K. Rothenberg, and O. Aalami. "Clinical validation of smartphone-based activity tracking in peripheral artery disease patients". In: *NPJ digital medicine* 1.1 (2018), pp. 1–8.

[30] A. J. Aviv, B. Sapp, M. Blaze, and J. M. Smith. "Practicality of Accelerometer Side Channels on Smartphones". In: ACSAC '12. Association for Computing Machinery, 2012.

[31] M. Bachlin, M. Plotnik, D. Roggen, I. Maidan, J. M. Hausdorff, N. Giladi, and G. Troster. "Wearable Assistant for Parkinson's Disease Patients With the Freezing of Gait Symptom". In: *IEEE Transactions on Information Technology in Biomedicine* 14.2 (2010), pp. 436–446.

[32] X. Bai, J. Yin, and Y.-P. Wang. "Sensor Guardian: prevent privacy inference on Android sensors". In: *Eurasip Journal on Information Security* 2017 (June 2017).

[33] M. Banky, R. A. Clark, B. F. Mentiplay, J. H. Olver, M. B. Kahn, and G. Williams. "Toward accurate clinical spasticity assessment: Validation of movement speed and joint angle assessments using Smartphones and camera tracking". In: *Archives of physical medicine and rehabilitation* 100.8 (2019), pp. 1482–1491.

[34] R. Bardenet, J. Flamant, and P. Chainais. *On the zeros of the spectrogram of white noise*. 2017. arXiv: 1708.00082 [stat.ME].

[35] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar. "Can Machine Learning Be Secure?" In: *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security*. ASIACCS. Association for Computing Machinery, 2006, 16–25.

[36] A. Bellet, R. Guerraoui, M. Taziki, and M. Tommasi. "Fast and Differentially Private Algorithms for Decentralized Collaborative Machine Learning". In: *CoRR* abs/1705.08435 (2017).

[37] C. BenAbdelkader, R. Cutler, and L. Davis. "Stride and cadence as a biometric in automatic person identification and verification". In: *IEEE FGR*. 2002, pp. 372–377.

[38] E. Bergamini, M. Iosa, V. Belluscio, G. Morone, M. Tramontano, and G. Vannozzi. "Multi-sensor assessment of dynamic balance during gait in patients with subacute stroke". In: *Journal of biomechanics* 61 (2017), pp. 208–215.

[39] D. J. Berndt and J. Clifford. "Using Dynamic Time Warping to Find Patterns in Time Series". In: *AAAIWS*. 1994, 359–370.

[40] S. D. Bersch, D. Azzi, R. Khusainov, I. E. Achumba, and J. Ries. "Sensor data acquisition and processing parameters for human activity classification". In: *Sensors* 14.3 (2014), pp. 4239–4270.

[41] L. Bloch. "Big assureur is watching you - alerte n°21". In: *Responsabilité civile et assurances* (2016).

[42] H. Bojinov, Y. Michalevsky, G. Nakibly, and D. Boneh. "Mobile Device Identification via Sensor Fingerprinting". In: *CoRR* abs/1408.1416 (2014).

[43] P. Bonato. "Advances in wearable technology and its medical applications". In: *2010 annual international conference of the IEEE engineering in medicine and biology*. IEEE. 2010, pp. 2021–2024.

[44] R. Bost, R. A. Popa, S. Tu, and S. Goldwasser. "Machine Learning Classification over Encrypted Data". In: *IACR Cryptology ePrint Archive* 2014 (2014), p. 331.

[45] N. Botts, B. Thoms, A. Noamani, and T. A. Horan. "Cloud Computing Architectures for the Underserved: Public Health Cyberinfrastructures through a Network of HealthATMs". In: *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*. HICSS '10. IEEE Computer Society, 2010, 1–10.

[46] L. Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[47] *Brightself application*. Health Technology Design research team. URL: https://htd.scss.tcd.ie/brightself/ (visited on 07/19/2021).

[48] C. F. Brinkløv, I. K. Thorsen, K. Karstoft, C. Brøns, L. Valentiner, H. Langberg, A. A. Vaag, J. S. Nielsen, B. K. Pedersen, and M. Ried-Larsen. "Criterion validity and reliability of a smartphone delivered sub-maximal fitness test for people with type 2 diabetes". In: *BMC Sports Science, Medicine and Rehabilitation* 8.1 (2016), pp. 1–9.

[49] J. Bushberg, J. Seibert, E. Leidholdt, and J. Boone. *The Essential Physics of Medical Imaging*. Wolters Kluwer Health, 2011, p. 280.

[50] D. Bzdok, N. Altman, and M. Krzywinski. *Points of significance: statistics versus machine learning*. 2018.

[51] L. Cai and H. Chen. "TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion". In: *HotSec*. 2011.

[52] N. A. Capela, E. D. Lemaire, and N. Baddour. "Novel algorithm for a smartphone-based 6-minute walk test application: algorithm, application development, and evaluation". In: *Journal of neuroengineering and rehabilitation* 12.1 (2015), pp. 1–13.

[53] N. Carlini, C. Liu, J. Kos, Ú. Erlingsson, and D. Song. "The Secret Sharer: Measuring Unintended Neural Network Memorization & Extracting Secrets". In: *CoRR* (2018).

[54] I. Carpinella, E. Gervasoni, D. Anastasi, T. Lencioni, D. Cattaneo, and M. Ferrarin. "Instrumental assessment of stair ascent in people with multiple sclerosis, stroke, and Parkinson's disease: a wearable-sensor-based approach". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26.12 (2018), pp. 2324–2332.

[55] S. Chakraborty, K. R. Raghavan, M. P. Johnson, and M. B. Srivastava. "A Framework for Context-Aware Privacy of Sensor Data on Mobile Systems". In: *HotMobile*. 2013.

[56] H. H. Chang, P. B. Chou, and S. Ramakrishnan. "An Ecosystem Approach for Healthcare Services Cloud". In: *2009 IEEE International Conference on e-Business Engineering*. 2009, pp. 608–612.

[57] J. Chen, J. Konrad, and P. Ishwar. *VGAN-Based Image Representation Learning for Privacy-Preserving Facial Expression Recognition*. 2018.

[58] S.-Y. Chen and C. J. Winstein. "A systematic review of voluntary arm recovery in hemiparetic stroke: critical predictors for meaningful outcomes using the international classification of functioning, disability, and health". In: *Journal of neurologic physical therapy : JNPT* 33.1 (2009), 2—13. ISSN: 1557-0576.

[59] Q. Cheng, J. Juen, S. Bellam, N. Fulara, D. Close, J. C. Silverstein, and B. Schatz. "Predicting pulmonary function from phone sensors". In: *Telemedicine and e-Health* 23.11 (2017), pp. 913–919.

[60] I. Y. Cheong, S. Y. An, W. C. Cha, M. Y. Rha, S. T. Kim, D. K. Chang, and J. H. Hwang. "Efficacy of mobile health care application and wearable device in improvement of physical performance in colorectal cancer patients undergoing chemotherapy". In: *Clinical colorectal cancer* 17.2 (2018), e353–e362.

[61] I. Chillotti, N. Gama, M. Georgieva, and M. Izabachène. "Faster Fully Homomorphic Encryption: Bootstrapping in Less Than 0.1 Seconds". In: *ASIACRYPT*. 2016, pp. 3–33.

[62] Y.-L. Chiu, Y.-J. Tsai, C.-H. Lin, Y.-R. Hou, and W.-H. Sung. "Evaluation of a smartphone-based assessment system in subjects with chronic ankle instability". In: *Computer methods and programs in biomedicine* 139 (2017), pp. 191–195.

[63] M Compagnat, S Mandigout, C. Batcho, N Vuillerme, J. Salle, R David, and J. Daviet. "Validity of wearable actimeter computation of total energy expenditure during walking in post-stroke individuals". In: *Annals of physical and rehabilitation medicine* 63.3 (2020), pp. 209–215.

[64] M. Compagnat, C. S. Batcho, R. David, N. Vuillerme, J. Y. Salle, J. C. Daviet, and S. Mandigout. "Validity of the walked distance estimated by wearable devices in stroke individuals". In: *Sensors* 19.11 (2019), p. 2497.

[65] S. Corporation. *Internet Security Threat Report*. 2017.

[66] K. Crager, A. Maiti, M. Jadliwala, and J. He. "Information Leakage through Mobile Motion Sensors: User Awareness and Concerns". In: 2017.

[67] M. Crawshaw. "Multi-Task Learning with Deep Neural Networks: A Survey". In: *CoRR* abs/2009.09796 (2020).

[68] C. Cui, G.-B. Bian, Z.-G. Hou, J. Zhao, G. Su, H. Zhou, L. Peng, and W. Wang. "Simultaneous Recognition and Assessment of Post-Stroke Hemiparetic Gait by Fusing Kinematic, Kinetic and Electrophysiological Data". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* PP (Mar. 2018), pp. 1–1.

[69] J. Dai, J. Teng, X. Bai, Z. Shen, and D. Xuan. "Mobile phone based drunk driving detection". In: *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. 2010.

[70] A. Das, N. Borisov, and M. Caesar. "Tracking Mobile Web Users Through Motion Sensors: Attacks and Defenses". In: Jan. 2016.

[71] P. DasMahapatra, E. Chiauzzi, R. Bhalerao, and J. Rhodes. "Free-living physical activity monitoring in adult US patients with multiple sclerosis using a consumer wearable device". In: *Digital biomarkers* 2.1 (2018), pp. 47–63.

[72] E. Davarcı, B. Soysal, I. Erguler, O. Sabri, O. Aydin, E. Dincer, and E. Anarim. "Age Group Detection Using Smartphone Motion Sensors". In: Sept. 2017.

[73] M. B. Del Rosario, N. H. Lovell, J. Fildes, K. Holgate, J. Yu, C. Ferry, G. Schreier, S.-Y. Ooi, and S. J. Redmond. "Evaluation of an mHealth-based adjunct to outpatient cardiac rehabilitation". In: *IEEE journal of biomedical and health informatics* 22.6 (2017), pp. 1938–1948.

[74] A. Derungs, C. Schuster-Amft, and O. Amft. "Longitudinal walking analysis in hemiparetic patients using wearable motion sensors: Is there convergence between body sides?" In: *Frontiers in bioengineering and biotechnology* 6 (2018), p. 57.

[75] S. Dey, N. Roy, W. Xu, R. Choudhury, and S. Nelakuditi. "AccelPrint: Imperfections of Accelerometers Make Smartphones Trackable". In: Jan. 2014.

[76] B. H. Dobkin, X. Xu, M. Batalin, S. Thomas, and W. Kaiser. "Reliability and validity of bilateral ankle accelerometer algorithms for activity recognition and walking speed after stroke". In: *Stroke* 42.8 (2011), pp. 2246–2250.

[77] P. Dohnalek, P. Gajdo, and T. Peterek. "Human Activity Recognition: Classifier Performance Evaluation on Multiple Datasets". In: *Journal of Vibroengineering* 16 (Jan. 2014).

[78] J. Domingo-Ferrer, D. Sánchez, and J. Soria-Comas. *Database Anonymization: Privacy Models, Data Utility, and Microaggregation-based Inter-model Connections*. Synthesis Lectures on Information Security, Privacy, & Trust. Morgan & Claypool Publishers, 2016.

[79] L. Ducas and D. Micciancio. "FHEW: Bootstrapping Homomorphic Encryption in Less Than a Second". In: *EUROCRYPT*. 2015, pp. 617–640.

[80] P. Düking, F. K. Fuss, H.-C. Holmberg, and B. Sperlich. "Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity". In: *JMIR mHealth and uHealth* 6.4 (2018), e102.

[81] C. Dwork. "Differential Privacy". In: *Automata, Languages and Programming*. Vol. 4052. 2006, pp. 1–12.

[82] P. Düking, S. Achtzehn, H.-C. Holmberg, and B. Sperlich. "Integrated Framework of Load Monitoring by a Combination of Smartphone Applications, Wearables and Point-of-Care Testing Provides Feedback that Allows Individual Responsive Adjustments to Activities of Daily Living". In: *Sensors* 18.5 (2018).

[83] P. Düking, A. Hotho, H.-C. Holmberg, F. K. Fuss, and B. Sperlich. "Comparison of Non-Invasive Individual Monitoring of the Training and Health of Athletes with Commercially Available Wearable Technologies". In: *Frontiers in Physiology* 7 (2016), p. 71.

[84] H. Edwards and A. Storkey. *Censoring Representations with an Adversary*. 2015.

[85] M. El-Gohary, D. Peterson, G. Gera, F. B. Horak, and J. M. Huisinga. "Validity of the instrumented push and release test to quantify postural responses in persons with multiple sclerosis". In: *Archives of physical medicine and rehabilitation* 98.7 (2017), pp. 1325–1331.

[86] G. Englebienne and H. Hung. "Mining for Motivation: Using a Single Wearable Accelerometer to Detect People's Interests". In: IMMPD '12. Association for Computing Machinery, 2012.

[87] M. K. Erb, D. R. Karlin, B. K. Ho, K. C. Thomas, F. Parisi, G. P. Vergara-Diaz, J.-F. Daneault, P. W. Wacnik, H. Zhang, T. Kangarloo, et al. "mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease". In: *NPJ digital medicine* 3.1 (2020), pp. 1–10.

[88] B. Ermis and A. T. Cemgil. *Differentially Private Dropout*. 2017.

[89] K. Evenson, M. Goto, and R. Furberg. "Systematic review of the validity and reliability of consumer-wearable activity trackers". In: *The international journal of behavioral nutrition and physical activity* 12 (2015), p. 159.

[90] H. Fang and Q. Qian. "Privacy Preserving Machine Learning with Homomorphic Encryption and Federated Learning". In: *Future Internet* 13.4 (2021).

[91] S. Fantozzi, M. Cortesi, A. Giovanardi, D. Borra, R. Di Michele, and G. Gatta. "Effect of walking speed during gait in water of healthy elderly". In: *Gait & Posture* 82 (2020), pp. 6–13.

[92] J. Farah, N. Baddour, and E. Lemaire. "Design, development, and evaluation of a local sensor-based gait phase recognition system using a logistic model decision tree for orthosis-control". In: *Journal of NeuroEngineering and Rehabilitation* 16 (Feb. 2019).

[93] M. Farooq, M. Waseem, A. Khairi, and P. Mazhar. "A Critical Analysis on the Security Concerns of Internet of Things (IoT)". In: *International Journal of Computer Applications* 111 (Feb. 2015), pp. 1–6.

[94] L. M. Feehan, J. Geldman, E. C. Sayre, C. Park, A. M. Ezzat, J. Y. Yoo, C. B. Hamilton, and L. C. Li. "Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data". In: *JMIR mHealth and uHealth* 6.8 (2018), e10527.

[95] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian. "Certifying and removing disparate impact". In: *SIGKDD*. 2015, pp. 259–268.

[96] A. Ferrari, P. Ginis, M. Hardegger, F. Casamassima, L. Rocchi, and L. Chiari. "A mobile Kalman-filter based solution for the real-time estimation of spatio-temporal gait parameters". In: *IEEE transactions on neural systems and rehabilitation engineering* 24.7 (2015), pp. 764–773.

[97] C. Ferreira, V. Guimarães, A. Santos, and I. Sousa. "Gamification of Stroke Rehabilitation Exercises Using a Smartphone". In: PervasiveHealth '14. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.

[98] *Fitbit application*. Fitbit. URL: https://www.fitbit.com/global/fr/products/services (visited on 07/19/2021).

[99] F. Flachenecker, H. Gaßner, J. Hannik, D.-H. Lee, P. Flachenecker, J. Winkler, B. Eskofier, R. A. Linker, and J. Klucken. "Objective sensor-based gait measures reflect motor impairment in multiple sclerosis patients: reliability and clinical validation of a wearable sensor device". In: *Multiple sclerosis and related disorders* 39 (2020), p. 101903.

[100] P. Flandrin. "Time–Frequency Filtering Based on Spectrogram Zeros". In: *IEEE Signal Processing Letters* 22.11 (2015), pp. 2137–2141.

[101] B. P. de Fontenay, J. S. Roy, B. Dubois, L. Bouyer, and J. F. Esculier. "Validating Commercial Wearable Sensors for Running Gait Parameters Estimation". In: *IEEE Sensors Journal* 20.14 (2020), pp. 7783–7791.

[102] *France investigates leak of almost 500,000 medical records, including HIV and fertility status*. News Wires. Feb. 25, 2021. URL: https://www.france24.com/en/europe/20210225-france-investigates-massive-leak-of-medical-records (visited on 07/18/2021).

[103] M. Fredrikson, S. Jha, and T. Ristenpart. "Model Inversion Attacks That Exploit Confidence Information and Basic Countermeasures". In: CCS. Association for Computing Machinery, 2015.

[104] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart. "Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing". In: SEC. USENIX Association, 2014.

[105] C. Frindel and D. Rousseau. "How Accurate Are Smartphone Accelerometers to Identify Intermittent Claudication?" In: Feb. 2018.

[106] S. Furtado, A. Godfrey, S. Del Din, L. Rochester, and C. Gerrand. "Are Accelerometer-based Functional Outcome Assessments Feasible and Valid After Treatment for Lower Extremity Sarcomas?" In: *Clinical Orthopaedics and Related Research®* 478.3 (2020), pp. 482–503.

[107] D. Gabor. "Theory of communication. Part 1: The analysis of information". English. In: *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering* 93 (26 1946), 429–441(12). ISSN: 0367-7540.

[108] M. Gadaleta, G. Cisotto, M. Rossi, R. Z. U. Rehman, L. Rochester, and S. Del Din. "Deep learning techniques for improving digital gait segmentation". In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 1834–1837.

[109] K. Ganju, Q. Wang, W. Yang, C. A. Gunter, and N. Borisov. "Property Inference Attacks on Fully Connected Neural Networks Using Permutation Invariant Representations". In: *ACM SIGSAC*. CCS '18 (2018), 619–633.

[110] E. Garcia-Ceja, V. Osmani, and O. Mayora-Ibarra. "Automatic Stress Detection in Working Environments From Smartphones' Accelerometer Data: A First Step". In: *IEEE Journal of Biomedical and Health Informatics* 20 (2016), pp. 1053–1060.

[111] *GDPR toolkit*. Commission Nationale de l'Informatique et des Libertés (CNIL). 2018. URL: https://www.cnil.fr/en/gdpr-toolkit (visited on 07/18/2021).

[112] J. Gehrke, E. Lui, and R. Pass. "Towards Privacy for Social Networks: A Zero-knowledge Based Definition of Privacy". In: *TCC*. 2011, pp. 432–449.

[113] C. Gentry, A. Sahai, and B. Waters. "Homomorphic Encryption from Learning with Errors: Conceptually-Simpler, Asymptotically-Faster, Attribute-Based". In: *CRYPTO*. 2013, pp. 75–92.

[114] S. M. Ghanem and I. A. Moursy. "Secure Multiparty Computation via Homomorphic Encryption Library". In: *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*. 2019, pp. 227–232.

[115] H. Ghasemzadeh, R. Jafari, and B. Prabhakaran. "A Body Sensor Network With Electromyogram and Inertial Sensors: Multimodal Interpretation of Muscular Activities". In: *IEEE Transactions on Information Technology in Biomedicine* 14.2 (2010), pp. 198–206.

[116] T. N. Gia, I. Tcarenko, V. K. Sarker, A. M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen. "IoT-based fall detection system with energy efficient sensor nodes". In: *2016 IEEE Nordic Circuits and Systems Conference (NORCAS)*. 2016, pp. 1–6.

[117] R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing. "CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by M. F. Balcan and K. Q. Weinberger. PMLR, 2016, pp. 201–210.

[118] O. Goldreich. "Cryptography and Cryptographic Protocols". In: *Distrib. Comput.* 16.2-3 (2003), pp. 177–199.

[119] O. Goldreich, S. Micali, and A. Wigderson. "A Completeness Theorem for Protocols with Honest Majority". In: *Conference Proceedings of the Annual ACM Symposium on Theory of Computing* (Jan. 1987).

[120] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. "Generative Adversarial Networks". In: (2014). eprint: 1406.2661.

[121] A. Goshvarpour and A. Goshvarpour. "Nonlinear Analysis of Human Gait Signals." In: *International Journal of Information Engineering & Electronic Business* 4.1 (2012).

[122] T. Graepel, K. Lauter, and M. Naehrig. "ML Confidential: Machine Learning on Encrypted Data". In: Nov. 2012, pp. 1–21.

[123] M. Gramaglia and M. Fiore. "Hiding mobile traffic fingerprints with GLOVE". In: *CoNEXT*. 2015, 26:1–26:13.

[124] B. Gregorutti, B. Michel, and P. Saint-Pierre. "Correlation and variable importance in random forests". In: *Statistics and Computing* 27.3 (2017), pp. 659–678.

[125] E. Grimpampi, V. Bonnet, A. Taviani, and C. Mazzà. "Estimate of lower trunk angles in pathological gaits using gyroscope data". In: *Gait & posture* 38.3 (2013), pp. 523–527.

[126] A. Gruenenfelder-Steiger, M. Katana, A. Martin, D. Aschwanden, J. Koska, Y. Kündig, E. Pfister-Lipp, and M. Allemand. "Physical Activity and Depressive Mood in the Daily Life of Older Adults". In: *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry* 30 (Aug. 2017), pp. 119–129.

[127] N. Gruschka and M. Jensen. "Attack Surfaces: A Taxonomy for Attacks on Cloud Services". In: *2010 IEEE 3rd International Conference on Cloud Computing*. 2010, pp. 276–279.

[128] *Hackers Selling More than 200 Million Stolen Data from Chinese Hotel Chain in Dark Web*. Balaji N, GBHackers On Security. Sept. 10, 2020. URL: https://gbhackers.com/hackers-selling-stolen-data/ (visited on 07/20/2021).

[129] J. Han, J. Pei, and M. Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[130] J. Han, E. Owusu, L. T. Nguyen, A. Perrig, and J. Zhang. "Accomplice: Location inference using accelerometers on smartphones". In: *COMSNETS*. 2012, pp. 1–9.

[131] R. M. Haralick, K. Shanmugam, and I. Dinstein. "Textural Features for Image Classification". In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-3.6 (1973), pp. 610–621.

[132] M. Hassanalieragh, A. Page, T. Soyata, G. Sharma, M. Aktas, G. Mateos, B. Kantarci, and S. Andreescu. "Health Monitoring and Management Using Internet-of-Things (IoT) Sensing with Cloud-Based Processing: Opportunities and Challenges". In: *2015 IEEE International Conference on Services Computing*. 2015, pp. 285–292.

[133] A. Henriksen, A.-S. Sand, T. Deraas, S. Grimsgaard, G. Hartvigsen, and L. Hopstock. "Succeeding with prolonged usage of consumer-based activity trackers in clinical studies: a mixed methods approach". In: *BMC Public Health* 20.1 (2020), pp. 1–14.

[134] J. Henriksen-Bulmer and S. Jeary. "Re-identification attacks—A systematic literature review". In: *International Journal of Information Management* 36.6, Part B (2016), pp. 1184 –1192.

[135] E. Hesamifard, H. Takabi, and M. Ghasemi. "Cryptodl: Deep neural networks over encrypted data". In: *arXiv preprint arXiv:1711.05189* (2017).

[136] E. Hesamifard, H. Takabi, and M. Ghasemi. "Deep Neural Networks Classification over Encrypted Data". In: CODASPY. 2019, pp. 97–108.

[137] S. Hidano, T. Murakami, S. Katsumata, S. Kiyomoto, and G. Hanaoka. "Model Inversion Attacks for Prediction Systems: Without Knowledge of Non-Sensitive Attributes". In: *15th Annual Conference on Privacy, Security and Trust (PST)*. 2017.

[138] B. Hitaj, G. Ateniese, and F. Pérez-Cruz. "Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning". In: *CoRR* abs/1702.07464 (2017).

[139] T. W. Hnat, E. Griffiths, R. Dawson, and K. Whitehouse. "Doorjamb: Unobtrusive Room-Level Tracking of People in Homes Using Doorway Sensors". In: SenSys '12. Association for Computing Machinery, 2012, 309–322.

[140] *Homomorphic Encryption for Arithmetic of Approximate Numbers*. https://github.com/snucrypto/HEAAN.

[141] P. R. C. House. "Mobile Health and Fitness Apps: What Are the Privacy Risks?" In: (2013).

[142] H. Hu, Z. Salcic, G. Dobbie, and X. Zhang. "Membership Inference Attacks on Machine Learning: A Survey". In: *CoRR* abs/2103.07853 (2021).

[143] J. Hua, Z. Shen, and S. Zhong. "We Can Track You if You Take the Metro: Tracking Metro Riders Using Accelerometers on Smartphones". In: *IEEE TIFS* 12.2 (2017), pp. 286–297.

[144] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar. "Adversarial Machine Learning". In: *Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence*. AISec. New York, NY, USA: Association for Computing Machinery, 2011, 43–58.

[145] T. Ilias, B. Filip, C. Radu, N. Dag, S. Marina, and M. Mevludin. "Using measurements from wearable sensors for automatic scoring of Parkinson's disease motor states: Results from 7 patients". In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2017, pp. 131–134.

[146] T. Isho, H. Tashiro, and S. Usuda. "Accelerometry-based gait characteristics evaluated using a smartphone and their association with fall risk in people with chronic stroke". In: *Journal of stroke and cerebrovascular diseases* 24.6 (2015), pp. 1305–1311.

[147] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller. "Deep learning for time series classification: a review". In: *Data Mining and Knowledge Discovery* 33.4 (2019), 917–963.

[148] J. F. Item-Glatthorn, N. C. Casartelli, J. Petrich-Munzinger, U. K. Munzinger, and N. A. Maffiuletti. "Validity of the intelligent device for energy expenditure and activity accelerometry system for quantitative gait analysis in patients with hip osteoarthritis". In: *Archives of physical medicine and rehabilitation* 93.11 (2012), pp. 2090–2093.

[149] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, and B. Li. "Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning". In: *2018 IEEE Symposium on Security and Privacy (SP)*. 2018, pp. 19–35.

[150] R. Jago, C. Anderson, T. Baranowski, and K. Watson. "Adolescent Patterns of Physical ActivityDifferences by Gender, Day, and Time of Day". In: *American journal of preventive medicine* 28 (June 2005), pp. 447–52.

[151] A. Jain and V. Kanhangad. "Investigating gender recognition in smartphones using accelerometer and gyroscope sensor readings". In: *2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT)*. 2016, pp. 597–602.

[152] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. Vol. 112. Springer, 2013.

[153] I.-Y. Jang, H. R. Kim, E. Lee, H.-W. Jung, H. Park, S.-H. Cheon, Y. S. Lee, and Y. R. Park. "Impact of a wearable device-based walking programs in rural older adults on physical activity and health outcomes: cohort study". In: *JMIR mHealth and uHealth* 6.11 (2018), e11335.

[154] C. Jayaraman, C. K. Mummidisetty, A. Mannix-Slobig, L. M. Koch, and A. Jayaraman. "Variables influencing wearable sensor outcome estimates in individuals with stroke and incomplete spinal cord injury: a pilot investigation validating two research grade sensors". In: *Journal of neuroengineering and rehabilitation* 15.1 (2018), pp. 1–18.

[155] W Johnson, O Onuma, M Owolabi, and S Sachdev. "Stroke: a global response is needed." In: 94,9 ().

[156] I. T. Jolliffe and J. Cadima. "Principal component analysis: a review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065 (2016).

[157] T. Jourdan, A. Boutet, and C. Frindel. "Toward privacy in IoT mobile devices for activity recognition". In: *MobiQuitous*. 2018, pp. 155–165.

[158] J. Juen, Q. Cheng, V. Prieto-Centurion, J. A. Krishnan, and B. Schatz. "Health monitors for chronic disease by gait analysis with mobile phones". In: *Telemedicine and e-Health* 20.11 (2014), pp. 1035–1041.

[159] J. Juen, Q. Cheng, and B. Schatz. "Towards a natural walking monitor for pulmonary patients using simple smart phones". In: *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*. 2014, pp. 53–62.

[160] J.-Y. Jung, M.-G. Chae, I. H. Jang, and H. Park. "A hybrid control method of an exoskeleton robot for intention-driven walking rehabilitation of stroke patients". In: *2012 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*. 2012, pp. 58–60.

[161] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. A. Bonawitz, Z. Charles, G. Cormode, R. Cummings, R. G. L. D'Oliveira, S. E. Rouayheb, D. Evans, J. Gardner, Z. Garrett, A. Gascón, B. Ghazi, P. B. Gibbons, M. Gruteser, Z. Harchaoui, C. He, L. He, Z. Huo, B. Hutchinson, J. Hsu, M. Jaggi, T. Javidi, G. Joshi, M. Khodak, J. Konečný, A. Korolova, F. Koushanfar, S. Koyejo, T. Lepoint, Y. Liu, P. Mittal, M. Mohri, R. Nock, A. Özgür, R. Pagh, M. Raykova, H. Qi, D. Ramage, R. Raskar, D. Song, W. Song, S. U. Stich, Z. Sun, A. T. Suresh, F. Tramèr, P. Vepakomma, J. Wang, L. Xiong, Z. Xu, Q. Yang, F. X. Yu, H. Yu, and S. Zhao. "Advances and Open Problems in Federated Learning". In: *CoRR* abs/1912.04977 (2019).

[162] G. Kaissis, A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, I. Lima, J. Mancuso, F. Jungmann, M.-M. Steinborn, A. Saleh, M. Makowski, D. Rueckert, and R. Braren. "End-to-end privacy preserving deep learning on multi-institutional medical imaging". In: *Nature Machine Intelligence* 3 (June 2021), pp. 1–12.

[163] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler. "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring". In: *TITB* 10.1 (2006), pp. 156–167.

[164] P. Kasnesis, C. Patrikakis, and I. Venieris. "PerceptionNet: A Deep Convolutional Neural Network for Late Sensor Fusion". In: *Proceedings of the 2018 Intelligent Systems Conference (IntelliSys) Volume 1* (Jan. 2019), pp. 101–119.

[165] M. Kearns and M. Li. "Learning in the Presence of Malicious Errors". In: *SIAM Journal on Computing* 22.4 (1993), pp. 807–837.

[166] M. Khatkar, K. Kumar, and B. Kumar. "An overview of distributed denial of service and internet of things in healthcare devices". In: *2020 Research, Innovation, Knowledge Management and Technology Application for Business Sustainability (INBUSH)*. 2020, pp. 44–48.

[167] L. Kidziński, S. Delp, and M. Schwartz. "Automatic real-time gait event detection in children using deep neural networks". In: *PLOS ONE* 14 (Jan. 2019), pp. 1–11.

[168] H. B. Kim, H. J. Lee, W. W. Lee, S. K. Kim, H. S. Jeon, H. Y. Park, C. W. Shin, W. J. Yi, B. Jeon, and K. S. Park. "Validation of freezing-of-gait monitoring using smartphone". In: *Telemedicine and e-Health* 24.11 (2018), pp. 899–907.

[169] J. Kim, N. Colabianchi, J. Wensman, and D. H. Gates. "Wearable Sensors Quantify Mobility in People With Lower Limb Amputation During Daily Life". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.6 (2020), pp. 1282–1291.

[170] P. Klasnja, S. Consolvo, T. Choudhury, R. Beckwith, and J. Hightower. "Exploring Privacy Concerns about Personal Sensing". In: *Pervasive Computing*. Ed. by H. Tokuda, M. Beigl, A. Friday, A. J. B. Brush, and Y. Tobe. Springer Berlin Heidelberg, 2009, pp. 176–183.

[171] A. F. R. Kleiner, I. Pacifici, A. Vagnini, F. Camerota, C. Celletti, F. Stocchi, M. F. De Pandis, and M. Galli. "Timed up and go evaluation with wearable devices: validation in Parkinson's disease". In: *Journal of bodywork and movement therapies* 22.2 (2018), pp. 390–395.

[172] D. Kobsar, J. M. Charlton, C. T. Tse, J.-F. Esculier, A. Graffos, N. M. Krowchuk, D. Thatcher, and M. A. Hunt. "Validity and reliability of wearable inertial sensors in healthy adult walking: A systematic review and meta-analysis". In: *Journal of neuroengineering and rehabilitation* 17 (2020), pp. 1–21.

[173] D. Kobsar, S. T. Osis, J. E. Boyd, B. A. Hettinga, and R. Ferber. "Wearable sensors to predict improvement following an exercise intervention in patients with knee osteoarthritis". In: *Journal of neuroengineering and rehabilitation* 14.1 (2017), pp. 1–10.

[174] N. Kostikis, D. Hristu-Varsakelis, M. Arnaoutoglou, and C. Kotsavasiloglou. "A Smartphone-Based Tool for Assessing Parkinsonian Hand Tremor". In: *IEEE Journal of Biomedical and Health Informatics* 19.6 (2015), pp. 1835–1842.

[175] V. Kovenko and I. Bogach. "A Comprehensive Study of Autoencoders' Applications Related to Images". In: *IT&I Workshops*. 2020.

[176] S. Kozey-Keadle, A. Libertine, K. Lyden, J. Staudenmayer, and P. S. Freedson. "Validation of wearable monitors for assessing sedentary behavior". In: *Medicine & Science in Sports & Exercise* 43.8 (2011), pp. 1561–1567.

[177] D. R. Krishnan, D. L. Quoc, P. Bhatotia, C. Fetzer, and R. Rodrigues. "IncApprox: A Data Analytics System for Incremental Approximate Computing". In: *WWW*. 2016, pp. 1133–1144.

[178] J. R. Kwapisz, G. M. Weiss, and S. A. Moore. "Cell phone-based biometric identification". In: *2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. 2010, pp. 1–7.

[179] D. T. H. Lai, R. K. Begg, and M. Palaniswami. "Computational Intelligence in Gait Research: A Perspective on Current Applications and Future Challenges". In: *IEEE Transactions on Information Technology in Biomedicine* 13.5 (2009), pp. 687–702.

[180] L. Lamberg. "Confidentiality and privacy of electronic medical records". In: *JAMA* 285.24 (2001), pp. 3075–3076.

[181] L. Lei, Y. Tan, K. Zheng, S. Liu, K. Zhang, and X. Shen. "Deep Reinforcement Learning for Autonomous Internet of Things: Model, Applications and Challenges". In: *IEEE Communications Surveys Tutorials* 22.3 (2020), pp. 1722–1760.

[182] D. Leightley, J. Darby, B. Li, J. S. McPhee, and M. H. Yap. "Human Activity Recognition for Physical Rehabilitation". In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*. 2013, pp. 261–266.

[183] D. Leightley, J. Mcphee, and M. H. Yap. "Automated Analysis and Quantification of Human Mobility Using a Depth Sensor". In: *IEEE Journal of Biomedical and Health Informatics* 21 (June 2016).

[184] J.-F. Lemay, A. Noamani, J. Unger, D. J. Houston, H. Rouhani, and K. E. Musselmann. "Using wearable sensors to characterize gait after spinal cord injury: evaluation of test–retest reliability and construct validity". In: *Spinal cord* (2020), pp. 1–9.

[185] R. Lemoyne and T. Mastroianni. "Implementation of a smartphone as a wireless accelerometer platform for quantifying hemiplegic gait disparity in a functionally autonomous context". In: *Journal of Mechanics in Medicine and Biology* 18.02 (2018), p. 1850005.

[186] N. Li, T. Li, and S. Venkatasubramanian. "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity". In: *ICDE*. 2007, pp. 106–115.

[187] N. Li, T. Li, and S. Venkatasubramanian. "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity". In: *2007 IEEE 23rd International Conference on Data Engineering*. 2007, pp. 106–115.

[188] S. Li, A. Ashok, Y. Zhang, C. Xu, J. Lindqvist, and M. Gruteser. "Whose move is it anyway? Authenticating smart wearable devices using unique head movement patterns". In: *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2016, pp. 1–9.

[189] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith. "Federated Learning: Challenges, Methods, and Future Directions". In: *CoRR* (2019).

[190] *Lifware spoon*. Liftware. URL: https://www.liftware.com/ (visited on 07/19/2021).

[191] F. Lipsmeier, K. I. Taylor, T. Kilchenmann, D. Wolf, A. Scotland, J. Schjodt-Eriksen, W.-Y. Cheng, I. Fernandez-Garcia, J. Siebourg-Polster, L. Jin, et al. "Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial". In: *Movement Disorders* 33.8 (2018), pp. 1287–1297.

[192] C. Liu, S. Chakraborty, and P. Mittal. *DEEProtect: Enabling Inference-based Access Control on Mobile Sensing Applications*. 2017.

[193] J. Loy-Benitez, S. Heo, and C. Yoo. "Soft sensor validation for monitoring and resilient control of sequential subway indoor air quality through memory-gated recurrent neural networks-based autoencoders". In: *Control Engineering Practice* 97 (2020), p. 104330.

[194] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee. "From local explanations to global understanding with explainable AI for trees". In: *Nature machine intelligence* 2.1 (2020), pp. 2522–5839.

[195] L. van der Maaten and A. Y. Hannun. "The Trade-Offs of Private Prediction". In: *CoRR* (2020).

[196] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. "l-diversity: Privacy Beyond k-anonymity". In: *ACM Transactions on Knowledge Discovery from Data* 1.1 (2007).

[197] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi. "Mobile sensor data anonymization". In: *IoTDI* (2019).

[198] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi. "Mobile Sensor Data Anonymization". In: *ACM IoTDI* (2019), pp. 49–58.

[199] M. Malekzadeh, R. G. Clegg, A. Cavallaro, and H. Haddadi. "Protecting Sensory Data Against Sensitive Inferences". In: *W-P2DS*. 2018, 2:1–2:6.

[200] S. Mallat. *A wavelet tour of signal processing 2nd Edition*. Academic Press, 1999.

[201] A. Mannini, S. Intille, M. Rosenberger, and A. Sabatini. "Activity Recognition Using a Single Accelerometer Placed at the Wrist or Ankle". In: *Medicine and science in sports and exercise* 45 (Apr. 2013).

[202] H. F. Maqbool, M. A. B. Husman, M. I. Awad, A. Abouhossein, N. Iqbal, and A. A. Dehghani-Sanij. "A real-time gait event detection for lower limb prosthesis control and evaluation". In: *IEEE transactions on neural systems and rehabilitation engineering* 25.9 (2016), pp. 1500–1509.

[203] M. J. Mathie, A. C. F. Coster, N. H. Lovell, B. G. Celler, S. R. Lord, and A. Tiedemann. "A pilot study of long-term monitoring of human movements in the home using accelerometry". In: *Journal of Telemedicine and Telecare* 10.3 (2004), pp. 144–151.

[204] A. Matic, V. Osmani, and O. Mayora. "Automatic Sensing of Speech Activity and Correlation with Mood Changes". In: *Pervasive and Mobile Sensing and Computing for Healthcare: Technological and Social Issues*. Ed. by S. C. Mukhopadhyay and O. A. Postolache. Springer Berlin Heidelberg, 2013, pp. 195–205.

[205] R. S. McGinnis, N. Mahadevan, Y. Moon, K. Seagers, N. Sheth, J. A. Wright Jr, S. DiCristofaro, I. Silva, E. Jortberg, M. Ceruolo, et al. "A machine learning approach for gait speed estimation using skin-mounted wearable sensors: From healthy controls to individuals with multiple sclerosis". In: *PloS one* 12.6 (2017), e0178366.

[206] H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas. "Federated Learning of Deep Networks using Model Averaging". In: *CoRR* (2016).

[207] S Mehrang, J. Pietilä, and I. Korhonen. "An Activity Recognition Framework Deploying the Random Forest Classifier and A Single Optical Heart Rate Monitoring and Triaxial Accelerometer Wrist-Band". In: *Sensors* 18.2 (2018), p. 613.

[208] C. Meisel, R. El Atrache, M. Jackson, S. Schubach, C. Ufongene, and T. Loddenkemper. "Machine learning from wristband sensor data for wearable, noninvasive seizure forecasting". In: *Epilepsia* 61.12 (2020), pp. 2653–2666.

[209] L. Melis, C. Song, E. D. Cristofaro, and V. Shmatikov. "Inference Attacks Against Collaborative Learning". In: *CoRR* abs/1805.04049 (2018).

[210] S. Menasria, J. Wang, and M. Lu. "The purpose driven privacy preservation for accelerometer-based activity recognition". In: *World Wide Web* 21 (2018), pp. 1773–1785.

[211] H. Menz, S. Lord, and R. Fitzpatrick. "Age-related differences in walking stability". In: *Age and ageing* 32 (Apr. 2003), pp. 137–42.

[212] I. Mileti, M. Germanotta, E. Di Sipio, I. Imbimbo, A. Pacilli, C. Erra, M. Petracca, S. Rossi, Z. Del Prete, A. R. Bentivoglio, et al. "Measuring gait quality in Parkinson's disease through real-time gait phase recognition". In: *Sensors* 18.3 (2018), p. 919.

[213] A. Mitrokotsa, M. Rieback, and A. Tanenbaum. "Classification of RFID Attacks". In: Jan. 2008, pp. 73–86.

[214] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Išgum. "Automatic Segmentation of MR Brain Images With a Convolutional Neural Network". In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1252–1261.

[215] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel. "Unique in the Crowd: The privacy bounds of human mobility". In: *Nature* 3 (2013).

[216] V. Morel. "Enhancing transparency and consent in the internet of things". Thèse de doctorat dirigée par Castelluccia, Claude et Le Métayer, Daniel Informatique Lyon 2020. PhD thesis. 2020.

[217] A. Moukadem, B. Zied, D. Ould-Abdeslamb, and A. Dieterlen. "A new optimized Stockwell transform applied on synthetic and real non-stationary signals". In: *Digital Signal Processing* 46 (2015), pp. 226–238.

[218] D. Munguía-Izquierdo, A. Santalla, and A. Legaz-Arrese. "Evaluation of a wearable body monitoring device during treadmill walking and jogging in patients with fibromyalgia syndrome". In: *Archives of physical medicine and rehabilitation* 93.1 (2012), pp. 115–122.

[219] *MyfitnessPal applications*. MyfitnessPal. URL: https://www.myfitnesspal.com/ (visited on 07/19/2021).

[220] A. Na and T. S. Buchanan. "Validating wearable sensors using self-reported instability among patients with knee osteoarthritis". In: *PM&R* 13.2 (2021), pp. 119–127.

[221] S. Na, L. Xumin, and G. Yong. "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm". In: *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*. 2010.

[222] B. Najafi, J. L. Helbostad, R. Moe-Nilssen, W. Zijlstra, and K. Aminian. "Does walking strategy in older people change as a function of walking distance?" In: *Gait & Posture* 29.2 (2009), pp. 261–266. ISSN: 0966-6362.

[223] S. Narain, T. D. Vo-Huu, K. Block, and G. Noubir. "Inferring User Routes and Locations Using Zero-Permission Mobile Sensors". In: *2016 IEEE Symposium on Security and Privacy (SP)*. 2016, pp. 397–413.

[224] M. Naseri, J. Hayes, and E. D. Cristofaro. "Toward Robustness and Privacy in Federated Learning: Experimenting with Local and Central Differential Privacy". In: *CoRR* (2020).

[225] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble. "Overview of use of decision tree algorithms in machine learning". In: *2011 IEEE Control and System Graduate Research Colloquium*. 2011, pp. 37–42.

[226] R. E. Navas, H. Le Bouder, N. Cuppens-Boulahia, F. Cuppens, and G. Papadopoulos. "Demo: Do not trust your neighbors! A small IoT platform illustrating a man-in-the-middle attack". In: *ADHOC-NOW: International Conference on Ad Hoc Networks and Wireless*. 2018, pp. 1–6.

[227] M. A. Newman, M. A. Hirsch, R. D. Peindl, N. A. Habet, T. J. Tsai, M. S. Runyon, T. Huynh, C. Phillips, N. Zheng, C. T. N. R. Group, et al. "Use of an instrumented dual-task timed up and go test in children with traumatic brain injury". In: *Gait & posture* 76 (2020), pp. 193–197.

[228] T. Nguyen Gia, V. K. Sarker, I. Tcarenko, A. M. Rahmani, T. Westerlund, P. Liljeberg, and H. Tenhunen. "Energy efficient wearable sensor node for IoT-based fall detection systems". In: *Microprocessors and Microsystems* 56 (2018), pp. 34–46. ISSN: 0141-9331.

[229] D. Nie, L. Wang, Y. Gao, and D. Shen. "Fully convolutional networks for multi-modality isointense infant brain image segmentation". In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. 2016, pp. 1342–1345.

[230] Nokia. "Threat Intelligence Report". In: (2020).

[231] B. T. Nukala, T. Nakano, A. Rodriguez, J. Tsay, J. Lopez, T. Q. Nguyen, S. Zupancic, and D. Y. Lie. "Real-time classification of patients with balance disorders vs. normal subjects using a low-cost small wireless wearable gait sensor". In: *Biosensors* 6.4 (2016), p. 58.

[232] S. J. Oh, M. Augustin, B. Schiele, and M. Fritz. *Towards Reverse-Engineering Black-Box Neural Networks*. 2018. arXiv: 1711.01768 [stat.ML].

[233] W. Oleszkiewicz, P. Kairouz, K. Piczak, R. Rajagopal, and T. Trzcinski. *Siamese Generative Adversarial Privatizer for Biometric Data*. 2018.

[234] A. Olsen and J. Torresen. "Smartphone accelerometer data used for detecting human emotions". In: Nov. 2016, pp. 410–415.

[235] F. J. Ordóñez and D. Roggen. "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition". In: *Sensors* 16.1 (2016), p. 115.

[236] E. Owusu, J. Han, S. Das, A. Perrig, and J. Zhang. "ACCessory: Password Inference Using Accelerometers on Smartphones". In: HotMobile. Association for Computing Machinery, 2012.

[237] J. O'Donoghue and J. Herbert. "Data Management within MHealth Environments: Patient Sensors, Mobile Devices, and Databases". In: 4.1 (2012).

[238] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng. "Recent progress on generative adversarial networks (GANs): A survey". In: *IEEE Access* 7 (2019), pp. 36322–36333.

[239] M. P. M. Parisot, B. Pejo, and D. Spagnuelo. "Property Inference Attacks on Convolutional Neural Networks: Influence and Implications of Target Model's Complexity". In: *CoRR* abs/2104.13061 (2021).

[240] N. Park, M. Mohammadi, K. Gorde, S. Jajodia, H. Park, and Y. Kim. "Data synthesis based on generative adversarial networks". In: *VLDB* 11.10 (2018), pp. 1071–1083.

[241] S. Patel, H.-S. Park, P. Bonato, L. Chan, and M. Rodgers. "A Review of Wearable Sensors and Systems with Application in Rehabilitation". In: *Journal of neuroengineering and rehabilitation* 9 (Apr. 2012), p. 21.

[242] J. Paulo, P. Peixoto, and U. Nunes. "ISR-AIWALKER: Robotic Walker for Intuitive and Safe Mobility Assistance and Gait Analysis". In: *IEEE Transactions on Human-Machine Systems* PP (Oct. 2017), pp. 1–13.

[243] J. M. Pavon, R. J. Sloane, C. F. Pieper, C. S. Colón-Emeric, H. J. Cohen, D. Gallagher, K. S. Hall, M. C. Morey, M. McCarty, and S. N. Hastings. "Accelerometer-Measured Hospital Physical Activity and Hospital-Acquired Disability in Older Adults". In: *Journal of the American Geriatrics Society* 68.2 (2020), pp. 261–265.

[244] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. "Scikit-learn: Machine learning in Python". In: *Journal of machine learning research* 12.Oct (2011), pp. 2825–2830.

[245] P. Pérez-Toro, J. Vásquez-Correa, T Arias-Vergara, E Nöth, and J. Orozco-Arroyave. "Nonlinear dynamics and Poincaré sections to model gait impairments in different stages of Parkinson's disease". In: *Nonlinear Dynamics* 100 (2020), pp. 3253–3276.

[246] A. Petit, T. Cerqueus, A. Boutet, S. B. Mokhtar, D. Coquil, L. Brunie, and H. Kosch. "SimAttack: private web search under fire". In: *Journal of Internet Services and Applications* 7.1 (2016), pp. 1–17.

[247] I. Poitras, F. Dupuis, M. Bielmann, A. Campeau-Lecours, C. Mercier, L. J. Bouyer, and J.-S. Roy. "Validity and reliability of wearable sensors for joint angle estimation: A systematic review". In: *Sensors* 19.7 (2019), p. 1555.

[248] V. Primault, A. Boutet, S. B. Mokhtar, and L. Brunie. "The Long Road to Computational Location Privacy: A Survey". In: *IEEE Communications Surveys Tutorials* 21.3 (2019), pp. 2772–2793.

[249] A. Pyae, M. Luimula, and J. Smed. "Understanding Stroke Patients' Motivation for Motivation-Driven Rehabilitative Game Design". In: July 2015.

[250] S. Qiu, L. Liu, H. Zhao, Z. Wang, and Y. Jiang. "MEMS inertial sensors based gait analysis for rehabilitation assessment via multi-sensor fusion". In: *Micromachines* 9.9 (2018), p. 442.

[251] W. Raghupathi and V. Raghupathi. "An Empirical Study of Chronic Diseases in the United States: A Visual Analytics Approach to Public Health". In: *International Journal of Environmental Research and Public Health* 15 (2018).

[252] P. Raknim and K.-c. Lan. "Gait monitoring for early neurological disorder detection using sensors in a smartphone: Validation and a case study of parkinsonism". In: *Telemedicine and e-Health* 22.1 (2016), pp. 75–81.

[253] N. Raval, A. Machanavajjhala, and J. Pan. "Olympus: Sensor Privacy through Utility Aware Obfuscation". In: *Proceedings on Privacy Enhancing Technologies* 2019.1 (2019).

[254] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. "Activity Recognition from Accelerometer Data". In: *Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3*. IAAI'05. Pittsburgh, Pennsylvania: AAAI Press, 2005, 1541–1546.

[255] F. Recher, O. Banos, C. D. Nikamp, L. Schaake, C. T. Baten, and J. H. Buurkc. "Optimizing Activity Recognition in Stroke Survivors for Wearable Exoskeletons". In: *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 2018, pp. 173–178.

[256] D. A. Revi, A. M. Alvarez, C. J. Walsh, S. M. De Rossi, and L. N. Awad. "Indirect measurement of anterior-posterior ground reaction forces using a minimal set of wearable inertial sensors: From healthy to hemiparetic walking". In: *Journal of neuroengineering and rehabilitation* 17.1 (2020), pp. 1–13.

[257] J. L. Reyes-Ortiz. *Smartphone-based human activity recognition*. Springer, 2015.

[258] M. Roberts, D. Mongeon, and F. Prince. "Biomechanical parameters for gait analysis: a systematic review of healthy human gait". In: *Phys Ther Rehabil* 4 (2017), p. 6.

[259] S. Rogan, R. de Bie, and E. D. de Bruin. "Sensor-based foot-mounted wearable system and pressure sensitive gait analysis". In: *Zeitschrift für Gerontologie und Geriatrie* 50.6 (2017), pp. 488–497.

[260] M. Romanelli, C. Palamidessi, and K. Chatzikokolakis. "Generating optimal privacy-protection mechanisms via machine learning". In: *arXiv preprint arXiv:1904.01059* (2019).

[261] M. Romei, M. Galli, F. Motta, M. Schwartz, and M. Crivellini. "Use of the normalcy index for the evaluation of gait pathology." In: *Gait & posture* 19 1 (2004), pp. 85–90.

[262] P. Rouget, A. Moukadem, A. Dieterlin, A. Boutet, and C. Frindel. "Anonymizing motion sensor data through time-frequency domain". In: *IEEE MLSP* (2021).

[263] D. S. Rubin, A. Dalton, A. Tank, M. Berkowitz, D. E. Arnolds, C. Liao, and R. M. Gerlach. "Development and pilot study of an iOS smartphone application for perioperative functional capacity assessment". In: *Anesthesia & Analgesia* 131.3 (2020), pp. 830–839.

[264] *Runkeeper application*. Asics. URL: runkeeper.com/cms/ (visited on 07/19/2021).

[265] *Runtastic application*. Adidas. URL: www.runtastic.com/ (visited on 07/19/2021).

[266] M. Rushanan, A. D. Rubin, D. F. Kune, and C. M. Swanson. "SoK: Security and Privacy in Implantable Medical Devices and Body Area Networks". In: *S&P*. 2014, pp. 524–539.

[267] H. Sadeghi. "Local or global asymmetry in gait of people without impairments." In: *Gait & posture* 17 3 (2003), pp. 197–204.

[268] A. Saha, A. Subramanya, and H. Pirsiavash. "Hidden Trigger Backdoor Attacks". In: *CoRR* abs/1910.00033 (2019).

[269] A. Salarian, F. B. Horak, C. Zampieri, P. Carlson-Kuhta, J. G. Nutt, and K. Aminian. "iTUG, a sensitive and reliable measure of mobility". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.3 (2010), pp. 303–310.

[270] N. Saleheen, A. A. Ali, S. M. Hossain, H. Sarker, S. Chatterjee, B. Marlin, E. Ertin, M. al'Absi, and S. Kumar. "PuffMarker: A Multi-Sensor Approach for Pinpointing the Timing of First Lapse in Smoking Cessation". In: UbiComp '15. Association for Computing Machinery, 2015.

[271] A. Salem, Y. Zhang, M. Humbert, M. Fritz, and M. Backes. "ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models". In: *CoRR* abs/1806.01246 (2018).

[272] S. Salman and X. Liu. "Overfitting Mechanism and Avoidance in Deep Neural Networks". In: *CoRR* abs/1901.06566 (2019).

[273] S. Sayyad. "Privacy Preserving Deep Learning using Secure Multiparty Computation". In: *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. 2020, pp. 139–142.

[274] E. S. Sazonov, G. Fulk, N. Sazonova, and S. Schuckers. "Automatic Recognition of postures and activities in stroke patients". In: *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2009, pp. 2200–2203.

[275] S Scalvini, D. Baratti, G. Assoni, M. Zanardini, L. Comini, and P. Bernocchi. *Information and communication technology in chronic diseases: a patient's opportunity*. 2014.

[276] F. Schäfer and A. Anandkumar. "Competitive gradient descent". In: *Advances in Neural Information Processing Systems*. 2019, pp. 7623–7633.

[277] D Schließmann, M Nisser, C Schuld, T Gladow, S Derlien, L Heutehaus, N Weidner, U Smolenski, and R Rupp. "Trainer in a pocket-proof-of-concept of mobile, real-time, foot kinematics feedback for gait pattern normalization in individuals after stroke, incomplete spinal cord injury and elderly patients". In: *Journal of neuroengineering and rehabilitation* 15.1 (2018), pp. 1–15.

[278] L. Schutte, U. Narayanan, J. Stout, P. Selber, J. Gage, and M. Schwartz. "An index for quantifying deviations from normal gait". In: *Gait & Posture* 11.1 (2000), pp. 25–31.

[279] M. Schwenk, G. S. Grewal, D. Holloway, A. Muchna, L. Garland, and B. Najafi. "Interactive sensor-based balance training in older cancer patients with chemotherapy-induced peripheral neuropathy: a randomized controlled trial". In: *Gerontology* 62.5 (2016), pp. 553–563.

[280] M. Schwenk, K. Hauer, T. Zieschang, S. Englert, J. Mohler, and B. Najafi. "Sensor-derived physical activity parameters can predict future falls in people with dementia". In: *Gerontology* 60.6 (2014), pp. 483–492.

[281] V. Seibert, R. Araújo, and R. McElligott. "Sensor Validation for Indoor Air Quality using Machine Learning". In: *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional*. SBC. 2020, pp. 730–739.

[282] E. Sejdic, I. Djurovic, and J. Jiang. "A Window Width Optimized S-Transform". In: *EURASIP J. Adv. Signal Process* (2008).

[283] S. Shema-Shiratzky, I. Hillel, A. Mirelman, K. Regev, K. L. Hsieh, A. Karni, H. Devos, J. J. Sosnoff, and J. M. Hausdorff. "A wearable sensor identifies alterations in community ambulation in multiple sclerosis: contributors to real-world gait quality and physical activity". In: *Journal of neurology* (2020), pp. 1–10.

[284] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga. "Fusion of Smartphone Motion Sensors for Physical Activity Recognition". In: *Sensors* 14.6 (2014), 10146–10176. ISSN: 1424-8220.

[285] R. Shokri, M. Stronati, C. Song, and V. Shmatikov. "Membership Inference Attacks Against Machine Learning Models". In: *IEEE SP* (2017), pp. 3–18.

[286] P. Singh, N. Juneja, and S. Kapoor. "Using Mobile Phone Sensors to Detect Driving Behavior". In: ACM DEV. Association for Computing Machinery, 2013.

[287] C. Song, T. Ristenpart, and V. Shmatikov. "Machine Learning Models that Remember Too Much". In: *CoRR* abs/1709.07886 (2017).

[288] S. SO'Dea. *Forecast Number of IoT Connected Objects Worldwide from 2018 to 2025, by Type*. 2018.

[289] F. Sposaro, J. Danielson, and G. Tyson. "iWander: An Android application for dementia patients". In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. 2010, pp. 3875–3878.

[290] S. Sprager and M. B. Juric. "Inertial sensor-based gait recognition: a review". In: *Sensors* 15.9 (2015), pp. 22089–22127.

[291] G. Sprint, D. J. Cook, D. L. Weeks, and V. Borisov. "Predicting functional independence measure scores during rehabilitation with wearable inertial sensors". In: *IEEE Access* 3 (2015), pp. 1350–1366.

[292] M. R. Srilekha and M. D. Jayakumar. "A Secure Screen Lock System for Android Smart Phones using Accelerometer Sensor". In: 2015.

[293] N. Srivastava and R. Salakhutdinov. "Multimodal Learning with Deep Boltzmann Machines". In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'12. Curran Associates Inc., 2012, 2222–2230.

[294] L. Stanković. "A measure of some time–frequency distributions concentration". In: *Signal Processing* 81.3 (2001), pp. 621–631.

[295] D. Stiawan, Y. Idris, R. Malik, S. Nurmaini, N. Alsharif, and R. Budiarto. "Investigating Brute Force Attack Patterns in IoT Network". In: *Journal of Electrical and Computer Engineering* 2019 (Apr. 2019), pp. 1–13.

[296] R. Stockwell, L. Mansinha, and R. Lowe. "Localization of the complex spectrum: the S transform". In: *IEEE Transactions on Signal Processing* 44.4 (1996), pp. 998–1001. DOI: 10.1109/78.492555.

[297] D. S. Stokic, T. S. Horn, J. M. Ramshur, and J. W. Chow. "Agreement between temporospatial gait parameters of an electronic walkway and a motion capture system in healthy and chronic stroke populations". In: *American journal of physical medicine amp; rehabilitation* 88.6 (2009), 437—444.

[298] N. Straiton, M. Alharbi, A. Bauman, L. Neubeck, J. Gullick, R. Bhindi, and R. Gallagher. "The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: A systematic review". In: *Maturitas* 112 (2018), pp. 85–93.

[299] L. Sweeney. "k-Anonymity: A model for protecting privacy". In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.5 (2002), pp. 557–570.

[300] Q. Tang, D. J. Vidrine, E. Crowder, and S. S. Intille. "Automated Detection of Puffing and Smoking with Wrist Accelerometers". In: PervasiveHealth '14. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.

[301] Y. Tang and C. Ono. "Detecting Activities of Daily Living from Low Frequency Power Consumption Data". In: *MOBIQUITOUS*. 2016, pp. 38–46.

[302] P. Terrier, J. Le Carré, M.-L. Connaissa, B. Léger, and F. Luthi. "Monitoring of gait quality in patients with chronic pain of lower limbs". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.10 (2017), pp. 1843–1852.

[303] W. Teufl, B. Taetz, M. Miezal, M. Lorenz, J. Pietschmann, T. Jöllenbeck, M. Fröhlich, and G. Bleser. "Towards an inertial sensor-based wearable feedback system for patients after total hip arthroplasty: Validity and applicability for gait classification with gait kinematics-based features". In: *Sensors* 19.22 (2019), p. 5006.

[304] *TFHE: Fast Fully Homomorphic Encryption over the Torus*. https://tfhe.github.io/tfhe/.

[305] *The Council of State asks the Health Data Hub for additional guarantees to limit the risk of transfer to the United States*. CNIL. Oct. 16, 2020. URL: https://www.cnil.fr/en/council-state-asks-health-data-hub-additional-guarantees-limit-risk-transfer-united-states (visited on 07/18/2021).

[306] *The Health App*. Apple. URL: https://www.apple.com/ios/health/ (visited on 07/19/2021).

[307] *The problem with AI ethics*. James Vincent, The Verge. Apr. 3, 2019. URL: https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech (visited on 07/20/2021).

[308] E. Thomaz, I. Essa, and G. D. Abowd. "A Practical Approach for Recognizing Eating Moments with Wrist-Mounted Inertial Sensing". In: UbiComp '15. Association for Computing Machinery, 2015.

[309] D. Toninelli, R. Pinter, P. Pedraza, I. Andreadis, C. Antoun, J. P. Azevedo, A. Ballivián, M. Couper, W. Durbin, G. Loewe, A. Mavletova, C. Ochoa, R. Poynter, M. Revilla, and A. Slavec. *Mobile Research Methods - Opportunities and challenges of mobile research methodologies*. Sept. 2015.

[310] F. Tramèr, F. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart. "Stealing Machine Learning Models via Prediction APIs". In: *CoRR* abs/1609.02943 (2016).

[311] A. Tripathy, Y. Wang, and P. Ishwar. "Privacy-preserving adversarial networks". In: *Allerton*. IEEE. 2019, pp. 495–505.

[312] A. Truong, A. Walters, J. Goodsitt, K. Hines, C. B. Bruss, and R. Farivar. "Towards automated machine learning: Evaluation and comparison of AutoML approaches and tools". In: *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE. 2019, pp. 1471–1479.

[313] M. Ullrich, A. Küderle, J. Hannink, S. Del Din, H. Gaßner, F. Marxreiter, J. Klucken, B. M. Eskofier, and F. Kluge. "Detection of gait from continuous inertial sensor data using harmonic frequencies". In: *IEEE Journal of Biomedical and Health Informatics* 24.7 (2020), pp. 1869–1878.

[314] D. Ummels, E. Beekman, K. Theunissen, S. Braun, and A. J. Beurskens. "Counting steps in activities of daily living in people with a chronic disease using nine commercially available fitness trackers: Cross-sectional validity study". In: *JMIR mHealth and uHealth* 6.4 (2018), e70.

[315] A. Vadnerkar, S. Figueiredo, N. E. Mayo, and R. E. Kearney. "Design and validation of a biofeedback device to improve heel-to-toe gait in seniors". In: *IEEE journal of biomedical and health informatics* 22.1 (2017), pp. 140–146.

[316] R. Vaiana, T. Iuele, V. Astarita, M. V. Caruso, A. Tassitani, C. Zaffino, and V. Giofré. "Driving Behavior and Traffic Safety: An Acceleration-Based Safety Evaluation Procedure for Smartphones". In: *Modern Applied Science* 8 (Dec. 2014), pp. 88–96.

[317] G. Vavoulas, C. Chatzaki, T. Malliotakis, M. Pediaditis, and M. Tsiknakis. "The MobiAct Dataset: Recognition of Activities of Daily Living using Smartphones". In: *ICT4AWE* (2016).

[318] P. Vergara, E. Marín, J. Villar, V. González, and J. Sedano. "An IoT Platform for Epilepsy Monitoring and Supervising". In: *Journal of Sensors* 2017 (July 2017), pp. 1–18.

[319] A. Vienne, R. P. Barrois, S. Buffat, D. Ricard, and P.-P. Vidal. "Inertial sensors to assess gait quality in patients with neurological disorders: a systematic review of technical and analytical challenges". In: *Frontiers in psychology* 8 (2017), p. 817.

[320] R. Vilallonga, A. Lecube, J. M. Fort, M. A. Boleko, M. Hidalgo, and M. Armengol. "Internet of Things and bariatric surgery follow-up: Comparative study of standard and IoT follow-up". In: *Minimally Invasive Therapy & Allied Technologies* 22.5 (2013), pp. 304–311.

[321] B. Wang and N. Z. Gong. "Stealing Hyperparameters in Machine Learning". In: *CoRR* abs/1802.05351 (2018).

[322] C. Wang, R. Goel, M. Noun, R. K. Ghanta, and B. Najafi. "Wearable Sensor-Based Digital Biomarker to Estimate Chest Expansion During Sit-to-Stand Transitions–A Practical Tool to Improve Sternal Precautions in Patients Undergoing Median Sternotomy". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28.1 (2019), pp. 165–173.

[323] J. Wang, S. Liu, and Y. Li. "A Review of Differential Privacy in Individual Data Release". In: *International Journal of Distributed Sensor Networks* 2015 (Oct. 2015), pp. 1–18. DOI: 10.1155/2015/259682.

[324] N. Wang, L. Jiao, P. Wang, M. Dabaghchian, and K. Zeng. "Efficient Identity Spoofing Attack Detection for IoT in mm-Wave and Massive MIMO 5G Communication". In: *2018 IEEE Global Communications Conference (GLOBECOM)*. 2018, pp. 1–6.

[325] Y. Wang, X. Wu, and D. Hu. "Using Randomized Response for Differential Privacy Preserving Data Collection". In: *EDBT*. 2016.

[326] G. Weiss and J. Lockhart. "Identifying User Traits by Mining Smart Phone Accelerometer Data". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Jan. 2011).

[327] G. M. Weiss, J. W. Lockhart, T. T. Pulickal, P. T. McHugh, I. H. Ronan, and J. L. Timko. "Actitracker: A Smartphone-Based Activity Recognition System for Improving Health and Well-Being". In: *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. 2016.

[328] *Wellness Coach - MyHealth application*. Terraillon. URL: https://www.terraillon.com/en/wellness-coach (visited on 07/19/2021).

[329] *What is The Quantified Self?* Gary Wolf. Mar. 3, 2011. URL: https://quantifiedself.com/blog/what-is-the-quantified-self/ (visited on 07/18/2021).

[330] J. R. Williamson, A. Dumas, G. Ciccarelli, A. R. Hess, B. A. Telfer, and M. J. Buller. "Estimating load carriage from a body-worn accelerometer". In: *IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 2015.

[331] R. D. Willmann, G. Lanfermann, P. Saini, A. Timmermans, J. t. Vrugt, and S. Winter. "Home Stroke Rehabilitation for the Upper Limbs". In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2007, pp. 4015–4018.

[332] K. Wilson and R. Dishman. "Personality And Physical Activity: A Systematic Review And Meta-analysis". In: *Medicine & Science in Sports & Exercise* 46 (May 2014), p. 473.

[333] *Withings application*. Withings. URL: https://www.withings.com/fr/en/ (visited on 07/19/2021).

[334] D. Wood, N. Apthorpe, and N. Feamster. "Cleartext Data Transmissions in Consumer IoT Medical Devices". In: *IoT S&P*. 2017, pp. 7–12.

[335] S. Wüest, F. Masse, K. Aminian, R. Gonzenbach, and E. D. De Bruin. "Reliability and validity of the inertial sensor-based Timed" Up and Go" test in individuals affected by stroke." In: *Journal of Rehabilitation Research & Development* 53.5 (2016).

[336] Z. Xu and S. Zhu. "SemaDroid: A Privacy-Aware Sensor Management Framework for Smartphones". In: CODASPY '15. Association for Computing Machinery, 2015.

[337] M. Yampolskiy, P. Horvath, X. Koutsoukos, Y. Xue, and J. Sztipanovits. "Taxonomy for description of cross-domain attacks on CPS". In: Apr. 2013, pp. 135–142.

[338] G. Yang, J. Deng, G. Pang, H. Zhang, J. Li, B. Deng, Z. Pang, J. Xu, M. Jiang, P. Liljeberg, H. Xie, and H. Yang. "An IoT-Enabled Stroke Rehabilitation System Based on Smart Wearable Armband and Machine Learning". In: *IEEE Journal of Translational Engineering in Health and Medicine* 6 (2018), pp. 1–10.

[339] Z. Yang, J. Zhang, E.-C. Chang, and Z. Liang. "Neural Network Inversion in Adversarial Setting via Background Knowledge Alignment". In: CCS. Association for Computing Machinery, 2019.

[340] A. C.-C. Yao. "How to generate and exchange secrets". In: *27th Annual Symposium on Foundations of Computer Science (sfcs 1986)*. 1986, pp. 162–167.

[341] S. Yeom, M. Fredrikson, and S. Jha. "The Unintended Consequences of Overfitting: Training Data Inference Attacks". In: *CoRR* (2017).

[342] T. K. Yoo, S. Kim, S. Choi, D. Kim, and D. W. Kim. "Interpretation of movement during stair ascent for predicting severity and prognosis of knee osteoarthritis in elderly women using support vector machine". In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2013 (July 2013), pp. 192–196.

[343] T. Yu, E. Bagdasaryan, and V. Shmatikov. "Salvaging Federated Learning by Local Adaptation". In: *CoRR* (2020).

[344] M. Zampolini, E. Todeschini, M. Guitart, H. Hermens, S. Ilsbroukx, V. Macellari, R. Magni, M. Rogante, S. Marchese, M. Vollenbroek Hutten, and C. Giacomozzi. "Telerehabilitation: Present and future". In: *Annali dell'Istituto superiore di sanità* 44 (Jan. 2008), pp. 125–34.

[345] E. Zarepour, M. Hosseini, S. S. Kanhere, and A. Sowmya. "A context-based privacy preserving framework for wearable visual lifeloggers". In: *PerCom*. 2016, pp. 1–4.

[346] Y. Zhai, N. Nasseri, J. Pöttgen, E. Gezhelbash, C. Heesen, and J.-P. Stellmann. "Smartphone accelerometry: A smart and reliable measurement of real-life physical activity in multiple sclerosis and healthy individuals". In: *Frontiers in neurology* 11 (2020), p. 688.

[347] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. "Understanding deep learning requires rethinking generalization". In: *CoRR* abs/1611.03530 (2016).

[348] J. Zhang, P. V. Orlik, Z. Sahinoglu, A. F. Molisch, and P. Kinney. "UWB Systems for Wireless Sensor Networks". In: *Proceedings of the IEEE* 97.2 (2009), pp. 313–331.

[349] L. Zhang, P. H. Pathak, M. Wu, Y. Zhao, and P. Mohapatra. "AccelWord: Energy Efficient Hotword Detection through Accelerometer". In: MobiSys. Association for Computing Machinery, 2015.

[350] S. Zhang, M. Jr, W. Xiao, and C. K. Tham. "Detection of Activities by Wireless Sensors for Daily Life Surveillance: Eating and Drinking". In: *Sensors (Basel, Switzerland)* 9 (Mar. 2009), pp. 1499–517.

[351] W. Zhang and B. Qu. "Security Architecture of the Internet of Things Oriented to Perceptual Layer". In: 2013.

[352] Z. Zhang, Y. Song, L. Cui, X. Liu, and T. Zhu. "Emotion recognition based on customized smart bracelet with built-in accelerometer". In: *PeerJ* 4 (July 2016), e2258.

[353] W. Zhou and S. Piramuthu. "Security/privacy of wearable fitness tracking IoT devices". In: *2014 9th Iberian Conference on Information Systems and Technologies (CISTI)*. 2014, pp. 1–5.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : JOURDAN
(avec précision du nom de jeune fille, le cas échéant)

DATE de SOUTENANCE : 28/10/2021

Prénoms : Théo Romain Julien

TITRE : Privacy and Transparency in learning systems for healthcare

NATURE : Doctorat

Numéro d'ordre : AAAALYSEIXXXX

Ecole doctorale : Interdisciplinaire Sciences Santé

Spécialité : Ingénieurie biomédicale, biotechnologie

RESUME :
Avec le développement de l'Internet des objets (IdO), les smartphones et les capteurs sont désormais capables de fournir des informations sur l'activité de l'utilisateur et même sur sa physiologie. Cela a donc suscité un intérêt croissant de la part de la communauté scientifique, notamment dans le domaine de la e-santé avec des applications dans le suivi des patients en cours de rééducation pour offrir un suivi plus personnalisé. Cependant, outre le fait de guider le processus de rééducation, la production et la transmission de données IdO sont également exposées à des atteintes à la vie privée. En effet, la chaîne de traitement complexe de l'application IdO dans les soins de santé multiplie les risques de menaces sur la vie privée tout au long du cycle de vie des données IdO, comprenant la collecte, la transmission et le stockage, par un adversaire qui peut récupérer les données et ré-identifier ou révéler des informations sensibles des patients. Cette thèse s'articule autour des questions suivantes: Les données collectées sont-elles suffisamment protégées pour que personne ne puisse en abuser pour ré-identifier le propriétaire ou déduire des informations sensibles? Les données protégées sont-elles encore suffisamment précises pour les applications de soins de santé telles que la rééducation? Atteindre cet équilibre entre l'utilité des données et la protection de la vie privée est un défi important que nous étudions dans cette thèse sous différents angles. Plus précisément, la première partie se concentre sur le problème de l'anonymisation des données par le biais de la minimisation, tandis que la deuxième partie se concentre sur la prévention de l'inférence d'attributs sensibles par le biais d'une approche basée sur les Réseaux Génératifs Adversariaux pour assainir les données des capteurs et une approche exploitant les couches privées dans l'apprentissage fédéré.

MOTS-CLÉS : traitement du signal, reconnaissance d'activité humaine, prédiction, apprentissage automatique, deep learning, apprentissage fédéré, réseaux génératifs adversariaux

Laboratoire (s) de recherche : CREATIS

Directeur de thèse: Carole Frindel

Président de jury :

Composition du jury : Caroline Fossati (Rapporteure), Emmanuel Vincent (Rapporteur), Aurélien Bellet (Examinateur), Sonia Ben Mokhtar (Examinatrice), Alain Dieterlen (Examinateur), Carole Frindel (Directrice de thèse), Antoine Boutet (Co-directeur de thèse)