COLLEGE MATHSTIC
DOCTORAL BRETAGNE
BRETAGNE OCEANE

Université Bretagne Sud

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ BRETAGNE SUD

ÉCOLE DOCTORALE N° 644
Mathématiques et Sciences et Technologies de l'Information et de la Communication en Bretagne Océane
Spécialité : *« Informatique »*

Par

« Qiyang LI »

« Ensuring privacy protection in data outsourcing by mobile applications and IoT »

Thèse présentée et soutenue à « IMT Atlantique, Brest », le « 17 Décembre 2025 »
Unité de recherche : « Lab-STICC »

Rapporteurs avant soutenance :

Estelle PAWLOWSKI Maîtresse de conférences HDR, Ecole Nationale Supérieure d'Ingénieurs de Caen
Benjamin NGUYEN Professeur, INSA Centre Val de Loire

Composition du Jury :

Présidente :	Caroline FONTAINE	Directrice de Recherche CNRS, Université Paris-Saclay
Examinateurs :	Estelle PAWLOWSKI	Maîtresse de conférences HDR, Ecole Nationale Supérieure d'Ingénieurs de Caen
	Benjamin NGUYEN	Professeur, INSA Centre Val de Loire
	Caroline FONTAINE	Directrice de Recherche CNRS, Université Paris-Saclay
	Willy ALLEGRE	Phd. Directeur technique, Centre de Kerpape
Dir. de thèse :	Guy GOGNIAT	Professeur, Université Bretagne Sud
Co-dir. de thèse :	David ESPES	Professeur, Université Bretagne Occidentale
Encadrante :	Johanne VINCENT	Maîtresse de conférences, IMT Atlantique

Acknowledgement

I would first like to express my deepest gratitude to my supervisors, Prof. David Espes, Prof. Guy Gogniat and Dr. Johanne Vincent. Thank you for choosing me to join this PhD project and for guiding me throughout the past three years. The challenges and struggles I experienced during this journey came mostly from the work itself, never from our collaboration. Working with you has always been a pleasure. Our discussions, your patience, and your constant encouragement, especially during the moments when I felt discouraged or overwhelmed, have meant a great deal to me. I feel truly fortunate to have had you as my mentors, and I am deeply grateful for the support and kindness you have shown me. This has been, in many ways, a genuinely enjoyable period of my life.

I would also like to thank the members of my thesis jury. I am grateful to my rapporteurs, Estelle PAWLOWSKI and Benjamin NGUYEN, for accepting this role and for their valuable comments on my manuscript. I also wish to thank Caroline FONTAINE and Willy ALLEGRE for serving as examiners and for their support as members of my CSI throughout these years. I truly appreciated their time, constructive feedback, and the pleasant exchanges we shared during my PhD.

I would like to thank my family as well for supporting my decisions. At the very beginning, they were hesitant about me coming to France for a PhD, and I fully understand their concerns, especially as the Covid had not yet completely ended when I started. I am sincerely grateful to my mother for ultimately trusting me and supporting my choice, despite her worries.

My heartfelt thanks also go to all the wonderful, friendly, and caring colleagues I have met in the department. I truly appreciate the warm and relaxed atmosphere that we shared. I am especially grateful to Armelle and Coraline, whose kindness, attentiveness, and constant help behind the scenes made our daily work so much easier. Their support has been invaluable, and I feel lucky to have worked with such thoughtful and generous people.

This work was carried out within the Handicap Innovation Territoire project, supported by the French Government as part of France 2030 - Territoires d'innovation program for innovative territories, and was administered by the "Banque des Territoires". I

am sincerely grateful for this financial support, without which this research would not have been possible.

Finally, I want to express my deepest appreciation to everyone who has offered me warmth, support, encouragement, and joy during this past year, a year when I struggled emotionally and often felt lonely. To my beautiful friends and to my beloved mother: thank you for being there for me, for lifting me up, and for giving me strength. Your presence has made all the difference, and I am profoundly grateful to have you in my life.

Thank you all for accompanying me on this journey.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

Table of Contents

Introduction **1**

1 Identifying and Classifying Privacy Concerns in E-Health and Corresponding Countermeasures **11**

1.1 Attacks . . . 11

1.1.1 Threat . . . 12

1.1.2 Taxonomy of privacy attacks . . . 13

1.1.3 Summary . . . 22

1.2 Privacy techniques . . . 23

1.2.1 Local . . . 24

1.2.2 Transmission/Transport . . . 35

1.2.3 Server/Cloud . . . 36

1.2.4 Summary . . . 40

1.3 Conclusion . . . 41

1.3.1 Local deployment constraints in real-world scenarios . . . 41

1.3.2 Robust privacy-preserving solutions . . . 42

1.3.3 Lack of datasets that include people with disabilities . . . 42

2 Attacker Models and the Datasets Used **43**

2.1 Attacker Models . . . 45

2.1.1 Supervised Learning Attacks . . . 45

2.1.2 Semi-Supervised Learning Attacks . . . 47

2.1.3 Unsupervised Learning Attacks . . . 51

2.2 Datasets . . . 53

2.2.1 HASC-IPSC . . . 54

2.2.2 WISDM . . . 56

2.3 Conclusion . . . 57

3 Privacy-Preserving Feature Selection in Local Environment 59
3.1 Multi-Objective Feature Selection . . . 63
3.1.1 Objectives of Privacy-Preserving Feature Selection . . . 64
3.1.2 Multi-objective Optimization for Feature Selection . . . 65
3.1.3 Classification Models . . . 68
3.2 Performance Evaluation . . . 68
3.2.1 Feature Extraction . . . 69
3.2.2 Experimental Setup . . . 70
3.3 Experimental Results . . . 71
3.3.1 Comparing different definitions of objectives . . . 71
3.3.2 Comparison of ways to select the optimal features . . . 76
3.3.3 Impact of different initial features . . . 78
3.4 Conclusion . . . 81

4 Robust Privacy-preserving Feature Selection in Human Activity Recognition Systems 83
4.1 Threat Model Instantiation and Analysis . . . 85
4.1.1 Unsupervised Learning Attacks . . . 86
4.2 Robust Privacy-Preserving Feature Selection . . . 87
4.2.1 Robust Privacy-Preserving Feature Selection . . . 87
4.3 Experimental Evaluation . . . 90
4.3.1 Feature Extraction . . . 92
4.3.2 Classifiers . . . 92
4.3.3 Experimental Setup . . . 92
4.4 Experimental Results . . . 94
4.4.1 Robustness Testing under Supervised Learning Attacks . . . 94
4.4.2 Robustness Testing under Semi-supervised Learning Attacks . . . 97
4.4.3 Robustness Testing under Unsupervised Learning Attacks . . . 98
4.4.4 Experimental Results of WISDM dataset . . . 99
4.5 Conclusion . . . 103

5 InclusiveHAR Dataset 105
5.1 Related Public Datasets . . . 106
5.2 Experimentation Setup . . . 108
5.3 Activities Setup . . . 110

5.4 Time-series Sensor Data . . . 112
5.5 Analysis of Time-Series Sensor Data . . . 114
5.5.1 Raw Data Comparison . . . 115
5.5.2 Statistical Information Analysis . . . 115
5.5.3 Trajectory Data Analysis . . . 117
5.6 Private Information Recognition Experiment . . . 118
5.6.1 Experiment with Supervised Learning Models . . . 119
5.6.2 Experiment with Unsupervised Learning Models . . . 121
5.7 Conclusion . . . 122

Conclusion **123**
Challenges in Implementing Local Privacy-Preserving Solutions . . . 125
HAR Data: Exploring the Depth and Breadth of Dataset Diversity . . . 126
Comprehensive Services for People with Disabilities: Balancing Care and Privacy Protection . . . 127
Legal Protection: Institutional Support for the Privacy of People with Disabilities 128

Bibliography **129**

Acronyms

AI Artificial Intelligence

API Application Programming Interface

ADL Activities of Daily Living

BLE Bluetooth Low Energy

CBIR Content-Based Image Retrieval

CNN Convolutional Neural Network

CDP Centralized Differential Privacy

DPDP Digital Personal Data Protection Act

DNS Domain Name System

DL Deep Learning

DNN Dense Neural Network

DP Differential Privacy

DDP Distributed Differential Privacy

DT Decision Tree

EHDS European Health Data Space

EMA Electromagnetic Analysis Attacks

ECG Electrocardiogram

EHR Electronic Health Record

EI Explicit Identifier

ESA Encode, Shuffle, Analyze architecture

FL Federated Learning

GDPR General Data Protection Regulation

GAN Generative Adversarial Network

HIPAA Health Insurance Portability and Accountability Act

HE Homomorphic Encryption

HDP Hybrid Differential Privacy

HAR Human Activity Recognition

HIT Handicap Innovation Territoire project

IoT Internet of Things

IoMT Internet of Medical Things

k-NN k-Nearest Neighbor

KMS Key Management System

LDP Local Differential Privacy

LSTM Long Short-Term Memory

LWC Lightweight Cryptography

LR Logistic Regression

MITM Man-in-the-Middle Attacks

ML Machine Learning

MIA Membership Inference Attacks

MI Model Inversion Attacks

MOEA Multi-Objective Evolutionary Algorithm

NSGA-III Non-dominated Sorting Genetic Algorithm III

NB Naive Bayes

OT Oblivious Transfer

PIPL Personal Information Protection Law of the People's Republic of China

PDP Personal Data Protection Bill

PAA Power Analysis Attacks

PHR Personal Health Record

PATE Private Aggregation of Teacher Ensembles

PIR Privacy Information Retrieval

PSI Private Set Intersection

QI Quasi-Identifier

RNN Recurrent Neural Network

RAE Replacement Autoencoder

RF Random Forest

SCA Side-Channel Attacks

SA Sensitive Attribute

SMC Secure Multi-party Computation

SVM Support Vector Machine

VAE Variational Autoencoder

VM Virtual Machine

XGB XGBoost

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

INTRODUCTION

Potential of IoT in e-Health Applications

Internet of Things (IoT) technology is rapidly becoming a key force in transforming the healthcare field, especially in e-health applications, which show unprecedented potential. By interconnecting various medical devices, smart sensors, wearables, and patient data platforms, IoT is breaking down the barriers of traditional healthcare models, enabling real-time data collection, remote monitoring and intervention, and personalized healthcare services. This seamless interconnectivity not only significantly improves healthcare efficiency and reduces operational costs, but more importantly, it greatly improves patient experience and health management, heralding the arrival of a new era of smarter, more efficient, and patient-centered healthcare [1]. For example, through smart wearable devices, key physiological indicators such as blood glucose and blood pressure of chronically ill patients can be continuously monitored and automatically uploaded to the cloud, allowing physicians to remotely view the data and make timely adjustments to the treatment plan, thus realizing effective care management. In addition, for the elderly or people living alone with limited mobility, smart home sensors can monitor their activity patterns and abnormalities, and send out immediate alerts in the event of falls and other emergencies, greatly improving the safety of home-based care.

Second, the IoT provides a powerful tool for disease prevention and early intervention. Through the continuous collection and analysis of daily health data, IoT systems are able to identify potential health risk trends and issue timely alerts to users or medical institutions. For example, smart bracelets and smartwatches can not only monitor heart rate, sleep quality and exercise, but also combine Artificial Intelligence (AI) techniques for in-depth analysis of these data to help users better understand their own health status and provide personalized health advice. This paradigm shift from "passive treatment" to "active prevention" will help reduce the incidence of disease, improve the health of the entire population, and ultimately reduce the burden on the entire healthcare system.

At a deeper level, the potential of IoT in e-health is also reflected in optimizing the operational efficiency of the entire healthcare system and promoting multi-party collab-

oration. It is capable of breaking down traditional information silos and realizing data interoperability and sharing among different healthcare institutions, departments and even patients' families, thus building a more integrated and responsive healthcare network. For example, first aid equipment connected through IoT can transmit patient vital signs data to hospital emergency rooms in advance, so that doctors can prepare for resuscitation before the patient arrives, which greatly shortens the rescue time. Meanwhile, in chronic disease management, family physicians, specialists and patients themselves can share health data and treatment plans through the same IoT platform, ensuring the consistency and effectiveness of treatment plans and truly realizing seamless patient-centered healthcare.

In summary, IoT technology, with its powerful connectivity and data processing capabilities, is profoundly changing the face of e-health, bringing a series of revolutionary applications that significantly enhance the efficiency, accessibility and personalization of healthcare services. Of particular interest is the crucial role that IoT technologies play in giving special groups, such as people with disabilities, independence and autonomy. With this in mind, the next section will explore in more detail how IoT-based e-health solutions can be specifically applied in the areas of care, health monitoring and assisted living for persons with disabilities, with an aim of enhancing their quality of life and independence.

Benefits and Challenges of e-Health for People with disabilities

In 2022 in mainland France, 14.5 million people aged 15 or over (28%) living at home have at least one severe functional limitation (e.g., vision or hearing problems despite correction), and 5.4 million (10%) are severely restricted in essential daily activities (getting in and out of bed, getting dressed, eating and drinking, etc.) [2]. Furthermore, the World Health Organization states that approximately 1.3 billion individuals worldwide are affected with disabilities [3]. It means that 16% of the world's population has special needs that differ from those of the general population. However, beyond typical patients and the elderly, the application of e-health in the lives of people with disabilities is often overlooked. This lack of consideration not only restricts progress in the development of assistive technologies but also underscores a more fundamental issue in inclusive research. Bridging this gap is crucial for the creation of targeted and effective solutions that enhance the quality of life for people with disabilities, promoting increased autonomy and

accessibility across diverse environments. In addition, they often encounter discrimination, such as work restrictions and high living costs (e.g., insurance and disability-specific services). Although the nature of the disability may vary, many of them need support to live healthy, safe, low-cost and independent lives.

However, their health needs and accessibility to health care have long been neglected. Traditional health systems often fail to adequately take into account the specific mobility, communication and cognitive impairments of people with disabilities, leading to difficulties in accessing timely and effective health services [4]. This neglect not only exacerbates their health inequalities, but also limits their opportunities to participate fully in society. In recent years, the rise of e-health has brought new hope for improving the healthcare experience for people with disabilities. Through various IoT and digital technologies, e-health is able to provide a range of convenient services. For example, through online consultation and remote diagnosis, patients with limited mobility can receive professional diagnosis and advice without having to go to the hospital; mobile health apps can provide personalized medication reminders and health monitoring functions to help people with disabilities better manage their own health; and virtual reality and augmented reality technologies can even be applied to rehabilitation training to provide a more immersive and interactive experience.

Despite the promising prospects of e-health, it still faces a series of challenges in practical applications. One notable issue is technology accessibility, i.e. not all persons with disabilities have equal access to and use of necessary electronic devices and Internet services. Economic conditions, living environment, and digital literacy could all pose barriers. In addition, the accessible design of the e-health platforms and devices is a key consideration. If systems do not meet accessibility standards, such as lack of screen reader compatibility or voice recognition functions, they can instead become new barriers that further exclude users with disabilities. Moreover, the level of acceptance and training of healthcare professionals in e-health is also critical. If healthcare professionals are unfamiliar with the remote consultation process or lack understanding of the special needs of persons with disabilities, the utility of e-health will be greatly diminished. Finally, data privacy and security are also a concern. The sensitivity of individuals with disabilities to personal health information requires a high level of security measures in e-health systems.

In summary, e-health paints a bright future for improving the accessibility of healthcare and quality of life for people with disabilities, but its successful implementation requires multi-party collaboration. We should not only focus on the popularization of the technol-

ogy, but also its inclusiveness. This includes ensuring that e-health solutions take into full account the diverse needs of patients from the very beginning of the design process, and that appropriate training and support are provided to them and healthcare professionals to bridge the digital skills gap. Furthermore, it is even more important to pay attention to data security and privacy, policy support, and many other aspects to ensure that e-health truly becomes an enabler for the health and well-being of people with disabilities, rather than a new barrier. Therefore, the next section will explore the importance and current status of privacy protection for people with disabilities.

Importance of Privacy Protection and Existing Privacy Concerns

The right to privacy is central to the dignity and autonomy of the individual, and for persons with disabilities, its importance is amplified as never before in today's context of smart healthcare and assisted living. With the proliferation of IoT devices, AI aids and telemedicine platforms, people with disabilities are enjoying unprecedented accessibility and independence. Smart wheelchairs can plan the best routes, wearable devices can monitor health indicators in real time, and AI voice assistants can aid in daily communication. Although these technological innovations have undoubtedly improved the quality of life for people with disabilities, at the same time, it also means that their lives are being digitized and datafied to an unprecedented extent. Protecting these sensitive data is not only about upholding their basic rights, but also about ensuring that while enjoying the convenience of technology, they are not deprived of their right to privacy and are not caught up in potential discrimination and victimization.

Currently, the main privacy concerns faced by people with disabilities in the smart healthcare and assisted living fields focus on the over-collection of sensitive personal data and the potential risk of misuse. As shown in Figure 1, smart medical devices can record physiological data ranging from heartbeat and breathing to sleep patterns; assisted living devices may monitor home activities and even record daily conversations between a person with a disability and his or her entourage through voice recognition; and a telerehabilitation application may store details of the person's rehabilitation progress and treatment. These devices can collect private information more frequently and in greater detail than traditional methods. Once these highly sensitive data are leaked, misused or accessed without authorization, not only may the sensitive attributes of people with

disabilities be maliciously exploited, leading to harassing advertisements and scams targeting persons with disabilities, but in more serious cases, they may be used for identity theft, discrimination in employment, or even, in extreme cases, the exposure of personal behavioral patterns that may lead to personal safety risks. As an example, criminals may exploit information about daily routines and mobility limitations exposed by smart devices to break into the homes of individuals with disabilities living alone or even carry out abductions.

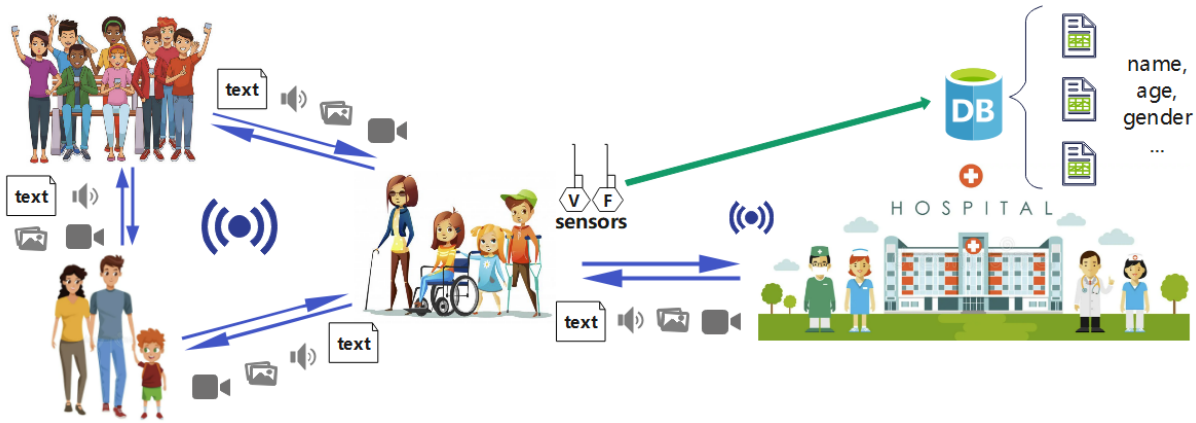


Figure 1 – Private data collection in e-health IoT systems for ambient assisted living

In addition, data security vulnerability and technological risks pose serious privacy threats. Smart healthcare and assisted living devices are often connected via the Internet, which means that they can be the target of cyberattacks. Once the servers or transmission links of these devices are breached, the medical records, home addresses, and even daily routines of people with disabilities may be compromised. At the same time, some data processors may have inadequate security protection during data storage and transmission, or may not anonymize data thoroughly enough, leaving "de-identified" data at risk of being re-identified, thus exposing personal privacy.

From a legal perspective, although many countries and regions have laws and regulations for personal data protection (e.g., GDPR in Europe or PIPL in China, as shown in Table 1), these laws apply to all citizens in principle, but they have certain limitations. They may not adequately consider the special needs of people with disabilities in terms of accessing information, understanding complex privacy policies, and expressing informed consent when confronted with the particularities of smart healthcare and assisted living scenarios. For example, for persons with disabilities who have cognitive or communication

difficulties, it is important to consider how to ensure that they truly understand and voluntarily give their consent for data usage. Existing laws may lack specific and actionable guidance in practice. As a result, the validity of their informed consent remains challenging. Furthermore, with technological advancements, new types of data (e.g., neural data) and new processing methods (e.g., AI-driven decision-making) are constantly emerging, and existing laws may lag behind or leave gaps in regulating these new issues.

Table 1 – Examples of privacy laws in different countries

Country	**Law**	**Summary and Features**
European Union	• General Data Protection Regulation (GDPR) [5] • European Health Data Space (EHDS) [6]	The GDPR contains very detailed provisions on General provisions, Principles, Independent supervisory authorities, Rights of data subjects, Controller and processor, Transfers of personal data to international organizations, etc. The EHDS aims to create a harmonized European digital health data ecosystem that facilitates the secure flow of health data within the European Union in order to provide better care for patients and support research and innovation.
China	Personal Information Protection Law of the People's Republic of China (PIPL) [7]	The most important features of PIPL are: Scope, key definitions and lawful processing of data; Individuals' rights and personal information handlers' duties; Personal information protection officer; Penalties and enforcement mechanisms; International data transfers.
U.S.	• California Privacy Rights Act • Virginia Consumer Data Protection Act • Colorado Privacy Act • Connecticut Data Privacy Act • Health Insurance Portability and Accountability Act (HIPAA)	Following California's lead, more than fifteen states (e.g., Colorado, Connecticut, and Virginia) have enacted or implemented new privacy protection statutes in recent years, and more states are sure to follow [8]. HIPAA, on the other hand, is a U.S. federal law designed primarily to protect the privacy and security of patients' health information and to regulate the handling of health care data and electronic transactions, while ensuring the portability of health insurance and combating fraud [9].
India	• Personal Data Protection (PDP) Bill 2019 (withdrawn in August 2022) • Digital Personal Data Protection (DPDP) Act 2023	The enactment of the PDP overhauled the personal data protection and regulatory regime in India. After its withdrawal, the DPDP came into effect in August 2023. It aims to enable personal data processing while recognizing individuals' rights and "the need to process personal data for lawful purposes", but it does not regulate non-personal data [10].

Therefore, to effectively address the privacy concerns of people with disabilities in smart healthcare and assisted living, a multi-level protection system needs to be constructed. This includes: requiring technology developers to incorporate the concept of

privacy protection at the product design stage; strengthening data security audits and regulation of smart devices and platforms; and making continuous improvements to the legal framework to adapt to technological development, as well as raising the awareness of privacy protection among persons with disabilities and their entourage. Only in this way can smart technologies truly become a blessing in the lives of people with disabilities, rather than a potential privacy trap.

Structure and Contributions of this Manuscript

The core objective of this thesis is to explore in depth and propose a comprehensive privacy protection solution to address the data privacy challenges faced by people with disabilities in the context of the growing popularity of smart healthcare and assisted living. We focus on identifying potential privacy leaks and risks of misuse that may arise from these emerging technologies (e.g., wearable devices) during data collection, processing, and sharing, and on building a multi-layered, sustainable, and actionable protection framework. Through this study, our aim is to provide a safer and more trustworthy smart living environment for people with disabilities, and to ensure that their personal data and dignity are fully respected and protected while enjoying the convenience of technology.

Chapter 1 provides an in-depth investigation of the multidimensional privacy threats that individuals with disabilities may face in smart healthcare and assisted living scenarios, along with the corresponding defense strategies. In order to construct a comprehensive and systematic analysis, this chapter meticulously categorizes and explores these threats and strategies in terms of the local data collection stage, the data transmission process, and the data storage and application stage in the cloud. This staged analysis is intended to lay a solid foundation for proposing targeted solutions in the subsequent chapters. This led to the following contribution.

> Qiyang Li, Johanne Vincent, David Espes, and Guy Gogniat, "Privacy challenges in e-health IoT systems for ambient assisted living," Under review.

Chapter 2 introduces the attack models considered in this thesis and the datasets used. The chapter details the various attack scenarios and their threats, as well as a comprehensive description of the sources and composition of the datasets used to ensure the transparency of the research and the credibility of the results.

Chapter 3 presents a locally applicable privacy-preserving feature selection scheme that aims to provide more comprehensive data protection for people with disabilities in smart healthcare and assisted living environments. The approach focuses on intelligent filtering and processing at the data source, i.e. the local device, to ensure that sensitive information has been effectively protected before it leaves the user's device, thus providing a more proactive and better guarantee of data security for persons with disabilities. This led to the following contribution.

> Qiyang Li, Johanne Vincent, David Espes, and Guy Gogniat, "Multi-objective feature selection for user privacy protection in human daily activity recognition," in 2025 21th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), Tuscany, Italy, June 9-11, 2025.

Chapter 4, as a further extension of the work in Chapter 3, proposes a more robust method of selecting privacy-preserving features. Building on the local protection in the previous chapter, this approach focuses on enhancing defense against various potential attacks, aiming to address the increasingly complex and covert privacy threats in smart healthcare and assisted living environments. By introducing advanced mechanisms, the work in this chapter strives to provide a more robust and comprehensive data security barrier for people with disabilities. This led to the following contribution.

> Qiyang Li, Johanne Vincent, David Espes, and Guy Gogniat, "Robust privacy-preserving feature selection in human activity recognition systems," Under review.

Considering the current lack of specific datasets that contain data on people with disabilities, this thesis, in collaboration with the Centre mutualiste de Kerpape [1], carefully designed and successfully collected a human activity recognition dataset including subjects with disabilities. Chapter 5 will provide a detailed description of the experimental environment, scenario setup, comprehensive data analysis, and the preliminary recognition results based on this dataset. This work lays a solid foundation for future data-driven research in the areas of assisted living and smart healthcare for people with disabilities. This led to the following contribution.

1. `https://hit-lorient.bzh/`

Qiyang Li, Johanne Vincent, David Espes, and Guy Gogniat, "InclusiveHAR: Open-access HAR Dataset including Sensor Data from People with Disabilities within Kerpape Living Lab," in the 18th International Symposium on Foundations & Practice of Security, Under review.

The final chapter summarizes the work done in this thesis on smart healthcare and assisted living privacy protection for people with disabilities. At the same time, this chapter examines the limitations of current localized privacy schemes in complex environments, and also reflects on the lack of diversity and information in existing datasets. In addition, it envisions how to better balance the functionality of smart applications with user privacy protection in the future, aiming to provide broader and more secure assistive services.

Chapter 1

Identifying and Classifying Privacy Concerns in E-Health and Corresponding Countermeasures

As a core driver of healthcare transformation in the digital age, e-health is profoundly changing the way healthcare services are delivered. From electronic medical records to telemedicine to wearable health devices, e-health technologies have greatly facilitated the circulation and sharing of healthcare-related information. However, behind these conveniences lies a huge test to the privacy and security of individuals' sensitive health data, especially for special groups such as people with disabilities. Due to the special sensitivity of health data, any privacy leakage may have serious impacts on individuals. Therefore, in-depth understanding of the privacy threats that exist in e-health and exploring effective countermeasure strategies are the current focus of both academia and industry. This chapter will survey the existing literature with the aim of presenting a comprehensive picture of the current status of privacy threats in e-health and their corresponding preventive measures.

1.1 Attacks

Personal and health information stored and transmitted by e-health systems is highly vulnerable to various malicious attacks. They not only originate from technical vulnerabilities in system design and implementation, but can also arise from mismanagement or malicious behavior. This section will first provide an overview of the common types of privacy threats in e-health environments, followed by an in-depth discussion of specific attacks, such as eavesdropping, inference attacks, etc., and a brief analysis of the harm they may cause.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

1.1.1 Threat

Confidentiality of private information is the main security property that need to be ensured. Depending on the nature and source of private information, privacy threats can be categorized as follows.

Identity disclosure. When an attacker confirms through various means that a record in a dataset belongs to an individual, identity disclosure occurs [11]. This can occur, for example, if an attacker is able to identify a person in a dataset that contains disability information. It represents the most serious disclosure because once it takes place, attackers can obtain sensitive information about specific individuals.

Attribute disclosure. In such attacks, the attackers aim to infer some sensitive information about a person [12], such as age, gender. Typically, attribute disclosure is a by-product of identity disclosure [13]. Attackers may exploit certain information attributes to make such inferences. If all individuals in a dataset share a sensitive attribute (e.g., leg disability), then assuming that the attacker knows someone is included, he will also know this information.

Background knowledge disclosure. When adversaries have some prior knowledge (e.g., age, gender) about a particular individual in a published dataset, they can infer other identifiable and sensitive information about him. Background knowledge originates from multiple sources, presenting countless possibilities and avenues, making it difficult to determine what background knowledge an attacker might possess [14]. For example, a person has access to information about his/her neighbor, such as zip code, gender, age, etc., then he/she may be able to gather additional information about that neighbor from public datasets (e.g., disability-related datasets).

Linkage disclosure. This disclosure occurs when attackers combine personal information from multiple sources to re-identify individuals or link records in otherwise anonymized datasets [13].

Membership disclosure. This occurs when adversaries can determine that a target individual is included in a dataset [15]. Obviously, if someone appears in a disability-related dataset, his/her disability status may be exposed even if the specific disability type cannot be inferred. In addition, it can serve as leverage to launch attribute disclosure [13].

Entourage disclosure. Existing research often overlooks the privacy risks faced by people with disabilities due to the actions of their entourage, such as family members, friends, or caregivers. However, their entourages may share information about them, either

inadvertently or even intentionally, leading to leakage of their privacy. This information is a priori harmless but can be cross-checked by malicious adversaries. It is precisely because of its stealth and flexibility that its threat level increases over time.

1.1.2 Taxonomy of privacy attacks

In this subsection, we propose our taxonomy of e-health privacy attacks faced by people with disabilities. These attacks are categorized based on the following three methods: 1) local, transmission or cloud; 2) plaintext or ciphertext; 3) active or passive. First, depending on the different phases, data may be attacked at different locations, and thus attacks can be categorized into local attacks (i.e. data collection), transmission attacks (i.e. data transfer), and server attacks (i.e. data storage and application). In addition, attacks can be categorized into plaintext and ciphertext based on the type of data they target. We refer to encrypted data using various encryption methods (e.g., symmetric & asymmetric encryption, Homomorphic Encryption (HE)) as ciphertext. Obviously, attack methods against plaintext are not applicable to ciphertext, and encryption can be used as a protection strategy for plaintext, but encrypted ciphertext still has the threat of privacy leakage, e.g., Side-Channel Attacks (SCA) [16]. Finally, the incorporation of active/passive categorization constitutes another contribution of our taxonomy, which is based on the activities of the attackers. Active attacks may affect the normal operations of the device, whereas passive attacks simply observe the behavior of the device during processing without interfering with it. The following subsections describe the attacks applied to three phases: local, transport and cloud. Within each category, we classify attacks as applying to plaintext, ciphertext or both, and discuss them from an active or passive perspective, respectively. With the help of this taxonomy, a better and clearer understanding of state-of-the-art e-Health privacy attacks can be achieved.

Local

This part will discuss attacks that can be executed locally, i.e., those targeting devices such as smartphones, tablets, and other equipment that collect and store personal data of people with disabilities. Most of the attacks here are ***physical attacks*** due to local distance limitations. They can be broadly categorized into two types: SCA and direct access attacks.

Local-Plaintext-Active. There are relatively few active attacks targeting plaintext,

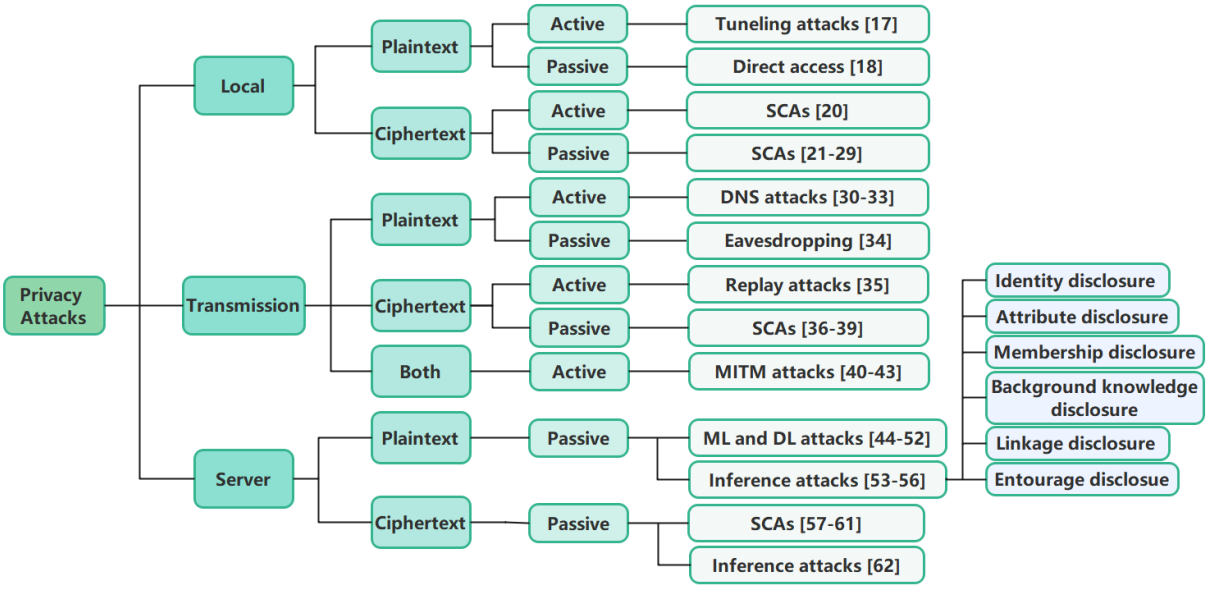


Figure 1.1 – Taxonomy of privacy attacks

as attackers in local scenarios tend to prefer covert passive attacks. Here, the focus is on a specific type of attack, namely ***tunneling attacks*** [17], which exploit protocol vulnerabilities to create network covert channels that bypass firewalls. Attackers can use these secret channels to steal sensitive information from local devices of people with disabilities. Such attacks are particularly noteworthy in the field of healthcare. DNS tunneling [17] is currently a popular method that may embed a person with a disability's private and medical data into DNS query requests, which are then transmitted directly to attacker-controlled servers via malicious DNS servers. Based on this, it could leak sensitive attributes or identity information.

Local-Plaintext-Passive. This part discusses passive ***physical attacks***. For local plaintext data, attackers may exploit ***direct access*** approach. This is a very simple and unobtrusive attack to obtain private information about people with disabilities by direct observation or direct physical access. For example, since many IoT sensor devices are deployed without protection in public areas, there exists physically direct access (e.g., node cloning [18]) to the sensor nodes themselves, resulting in the leakage of private medical data stored on the nodes. These attacks, which work on plaintext, not only threaten the sensitive attributes of people with disabilities, but also have the potential to cause identity disclosure.

Local-Ciphertext-Active. ***Side-Channel Attacks*** *(SCA)* are a class of physical attacks and assume that the data of people with disabilities are always leaked even when encrypted, thus allowing adversaries to exploit these leakages [19]. These attacks extract information from the cryptographic module by analyzing statistical data on various operational characteristics of the physical device [19]. In general, SCA that use input signals are active attacks, and those that use output signals are passive attacks. Since SCA techniques are applied to ciphertext, they primarily lead to attribute disclosure rather than directly threatening identity information. However, the information obtained through SCA can serve as a source of background knowledge leakage.

A particular ***active acoustic analysis attack***, SonarSnoop [20], can be remotely executed to infer private information. Although the SonarSnoop technique described in the article is applied to recovering smartphone unlock patterns, its underlying principle of using acoustic side-channel analysis of users' physical interactions could also pose a potential threat to medical and health data. For instance, if a doctor or patient enters sensitive health information on a touchscreen device, attackers could theoretically utilize similar techniques to exploit the device's acoustic system in order to steal this data, thereby disclosing the patient's sensitive attributes.

Local-Ciphertext-Passive. The remaining SCA are passive attacks. ***Power Analysis Attacks*** *(PAA)* are based on observing, measuring and analyzing the power consumption of cryptographic devices to steal sensitive inputs, e.g., recovering user ECG data from a fitness tracker [21]. ***Electromagnetic Analysis Attacks*** *(EMA)* utilize EM signals of different frequencies and amplitudes emitted when a device executes different commands to infer the internal operations performed and the data processed by the device, such as encryption keys [22], or images displayed on the screen [23]. Although [23] does not directly focus on the field of healthcare, the proposed attack method is generalizable and thus may also pose potential threats to healthcare-related devices. Since EM signals can be captured remotely, EMA can be executed over long distances using strong signal-receiving devices [23]. ***Acoustic analysis attacks*** exploit acoustic emissions generated during the operation of electronic devices to infer and recover secret information. For example, an attacker could reconstruct medical prescriptions printed by a printer [24], or analyze the minute vibrations of differential pressure sensors induced by sound waves (e.g., human voices) to reconstruct nearby conversations, potentially compromising sensitive private information within healthcare settings [25]. Additionally, attackers can exploit side channels, such as cache and power, to steal encryption keys [26]. Although

these attack techniques are complex and difficult to implement, they pose a potential threat to devices in healthcare and assisted living scenarios, as they can bypass conventional defenses and thus threaten personal health data.

Today, attackers can leverage unintentional information leaks generated by smartphone hardware or software to infer sensitive user data, which is known as ***smartphone-based SCA***. For example, Rehman et al. [27] employed the vibration generated by typing on a smartphone's soft keyboard, using hardware sensors to infer keystrokes, thereby stealing sensitive healthcare data. Whereas Spreitzer et al. [28] inferred user browsing behavior by accessing unprotected data usage statistics to obtain sensitive personal information such as health status, even when using an anonymous network. A novel ***timing analysis attack*** [29] can analyze the temporal differences generated during neural network computation to infer its classification results. It mainly targets neural networks on embedded devices, e.g., arrhythmia detection in health monitoring devices. Attackers can remotely launch these novel SCA without physical access to the device, additional tools, or specific permissions, so extra care should be taken to protect against such attacks.

However, it is worth mentioning that all traditional SCA require the attacker to be close to the device in order to access the statistics, and even EMA have distance limitations. In addition, they have requirements on the environment in which the data is measured, especially in terms of power and acoustic analysis. Therefore, these traditional SCA are less likely in the context of assisted living scenarios for people with disabilities, but their threat to the privacy of people with disabilities still cannot be ignored.

Transmission/Transport

This part discusses attacks that may occur during data transmission. Compared to local attacks, these attacks are more flexible, diverse, and widespread, as data passes through multiple points, enabling various attack vectors such as network traffic monitoring (even on encrypted networks) and exploitation of communication protocol vulnerabilities. According to our taxonomy, the discussion starts with attacks that target plaintext data.

Transport-Plaintext-Active. DNS tunneling has been described before, but since DNS is a communication protocol, more ***DNS attacks*** can be executed during the data transmission stage. ***DNS spoofing*** redirects communication traffic to malicious servers by tampering with the domain name resolution process, thus stealing sensitive identity and attribute information [30]. Such attacks pose a real-world threat in healthcare and assisted living environments. According to [31], numerous IoT devices rely on vulnerable

DNS implementations, which can be remotely exploited to cause data leakage, directly threatening the security and privacy of medical data. Furthermore, domain name queries of IoT devices are often highly distinctive, enabling attackers to easily identify and locate specific devices [32]. This further increases the risk of sensitive health data and lifestyle information being compromised by DNS spoofing. In smart robot applications for healthcare and assisted living, the integration of security services is regarded as a critical issue, where the integrity and trustworthiness of DNS-related communications form a vital foundation for system trust [33]. However, despite extensive research on DNS spoofing within general cybersecurity, systematic exploration targeting healthcare and assisted living scenarios remains relatively limited and is still in its infancy.

Transport-Plaintext-Passive. ***Eavesdropping*** is a passive Man-in-the-Middle (MITM) attack that only observes traffic passing through a specific network node without affecting the network or user behavior. [34] points out that some wearable BLE-enabled health monitoring devices lack effective end-to-end encryption during data transmission, exposing sensitive information over wireless links in plaintext or with weak protection. This allows attackers to directly access users' sensitive health data, such as heart rate. It could reveal sensitive identities, attributes, and background knowledge. However, this aspect needs to be further explored in the field of healthcare.

Transport-Ciphertext-Active. In assisted living scenarios, ***replay attacks*** can steal the privacy of people with disabilities by reusing information that has been exchanged during data transmission to trick devices into performing sensitive operations, such as uploading private data [35].

Transport-Ciphertext-Passive. Unlike local SCA, there are other ways to implement ***SCA*** during data transmission. ***Timing analysis attacks*** analyze the timing of special events or operations to uncover hidden information in signal transmission systems. As an example, attackers can exploit variations in cache response time to infer disease contained in user queries to large language models [36], e.g., when a query contains terms related to diabetes (such as "blood glucose" or "insulin"), the model may respond faster due to higher cache hits for related content; conversely, queries without such terms may take longer, potentially allowing attackers to infer that the user might have diabetes. Moreover, attackers can also utilize the physical side channels of intermittent devices, such as the turn-on/off patterns revealed by inter-packet arrival times, to infer user daily activities and routines, thereby potentially disclosing user privacy such as health status and lifestyle habits [37]. ***Traffic analysis attacks*** categorize traffic by observing

the metadata of encrypted network traffic, and the features used for statistical analysis include count, size (e.g., total bytes transferred), [38] etc. Typically, attackers combine traffic and timing analysis. From the traffic of healthcare websites, attackers can infer user medication/surgery/disease information [38]. Furthermore, adversaries can monitor the browsing behavior of people with disabilities and identify the web pages they visit by analyzing encrypted DNS traffic, thus inferring sensitive information about them, e.g., disability status [39]. Both SCA are useful for attackers who wish to obtain background knowledge from wireless networks.

Transport-Both. This part introduces attacks that may be applied to plaintext and ciphertext. They are categorized this way not because of their flexibility, but because not all data transmitted over a network is encrypted.

Transport-Both-Active. ***MITM*** Attacks occur when attackers control the communication channel between two endpoints without the victims' knowledge. In other words, they believe that their communication channel is secure. MITM attacks can eavesdrop (passively), modify, or intercept (actively) intended communication traffic [40]. In the medical domain, such attacks primarily occur in Bluetooth Low Energy (BLE) connections used by wearable and remote monitoring devices, as well as in the broader Internet of Medical Things (IoMT) environment. Melamed et al. [41] demonstrated how adversaries can intercept and manipulate health-related data , such as modifying exaggerated kilometrage; the authors warn that this technique could also target life-depending devices (e.g., diabetic/blood-sugar monitors). Similarly, Yurdagul et al. [42] show that physiological parameters, such as heart rate, are highly susceptible to be leaked during wireless transmission. Extending to IoMT systems, Salem et al. [43] reveal that attackers may replace genuine abnormal readings with "normal" values, thus suppressing alarms and causing patients' health status records to be exposed and manipulated. In general, these findings indicate that MITM attacks in healthcare could threaten patient privacy by leaking physiological data, treatment-related information, and identity attributes, posing serious challenges to the reliability and compliance of remote monitoring and care systems.

The attacks on network protocols discussed in this section, i.e., MITM and DNS-related attacks, have seen relatively limited research in the context of healthcare applications. However, due to their wide attack scope, not only the personal information of people with disabilities, e.g., identity, and disability status, but also e-health data can also be threatened by them. In addition, most passive eavesdropping attacks, replay attacks, and some active tampering attacks (i.e., some MITM attacks), can be defended against

by disabling plaintext transmission, using secure communication protocols, implementing mutual authentication, etc. Nevertheless, even with these measures, there are still threats such as SCA, so in the next privacy protection section, our study will focus on defense countermeasures against these specific threats.

Server/Cloud

This subsection discusses the privacy threats to people with disabilities when data is stored and processed on the server. It is assumed that the data of people with disabilities have been securely transferred to the cloud, but here they still face some new challenges.

Server-Plaintext-Passive. ***Machine Learning*** *(ML)* and ***Deep Learning*** *(DL)* are rapidly advancing, but as their applications expand and real-world deployments increase, privacy leakage risks are emerging. In recent years, an increasing number of AI applications have begun to process sensitive information (e.g., conversations between close individuals, medical image data), and therefore, researchers have begun to focus on these privacy exposure issues, which include potential exposure of identities and sensitive attributes, membership in the training data, and the inference of additional information from model outputs.

First, the most serious one is ***Identity Inference Attacks***. Wang et al. [44] proposed an mGAN-AI framework for federated learning to infer user-level privacy, which can simultaneously discriminate the category, reality, and client identity of the input samples, making it possible to reconstruct the user privacy. Then ***Membership Inference attacks*** *(MIA)* can detect whether a given sample belongs to the original training dataset of the target model [45]. For example, attackers can infer whether a patient has been hospitalized in a hospital and received a particular major procedure [45]. In healthcare, sharing synthetic data is considered as a method that can protect patient privacy; however, synthetic data generation models still face the threat of MIA [46]. Moreover, social media health data may also reveal the psychological states of associated users [47]. As for ***Reconstruction attacks***, also known as ***Model Inversion attacks*** *(MI)* or ***Attribute Inference attacks***, adversaries can infer sensitive features of the training data or original inputs from model outputs [48]. Unlike MIA, these attacks can reconstruct the data in the training set. It is commonly used by attackers in the medical field to reconstruct medical image, e.g., [49], and it could also reconstruct sensitive attributes such as disability status. In general, the more predictive a model is, the more vulnerable it is to such attacks [48]. Sometimes, in addition to sensitive features of the training data, attackers may also have

access to information about the dataset, such as statistical characteristics (e.g., disease distribution) that can be used to infer data owner preferences, i.e., ***Property Inference attacks***. Since the information is independent of the data and the training process, it is likely to be accidentally leaked, which is unavoidable. Well-generalized models also face such problems [50]. Imagine that attackers can obtain private information of a person with a disability, e.g., medical records, simply by accessing a personalized neural network model for natural language text, this is a ***Secret Extraction attack***. Carlini et al. [51] successfully implemented this attack and identified social security numbers and credit card numbers. Attackers could steal sensitive sequences from text prediction models used in chat apps, emails, conversational AI bots, etc., through targeted and continuous queries. With the popularity of these models, such as ChatGPT, this threat to privacy cannot be ignored. In addition, recent research [52] indicates that privacy risks arise not only from the model itself but also from its surrounding system components, such as training data filtering, input preprocessing, output filtering, and query control. These components may create new ***"privacy side channels"***, making it easier for attackers to infer or steal sensitive information. In healthcare or assisted living scenarios, this means that processes, e.g., deduplication of EHRs, input checks for remote consultations, or automatic filtering of care instructions could inadvertently expose patient privacy.

Beyond privacy leakage attacks in AI models, publicly released datasets can remain vulnerable to ***Inference attacks***, even after anonymization. Given the diversity and variability of these attacks, all types of disclosure discussed in Section 1.1.1 may occur; however, attribute inference and linkage disclosure represent the most salient risks. Among these anonymized data, graph data (see Figure 1.2) are special, because with this form, graph data can anonymize private information and keep only useful information (e.g., medical data), but the links between graph nodes still compromise privacy. Existing graph embedding algorithms typically do not consider privacy risks arising from nodes and their connections, which leads to ***Graph Embedding Inference attacks***. Attackers can infer sensitive attributes by analyzing node and vector representations learned by graph embedding algorithms [53]. ***Homogeneity attacks*** are methods of obtaining sensitive attributes in k-anonymous data tables by comparing the data in the columns of the table and identifying common features [54], see Figure 1.3, Tom suffers from heart disease, since everyone in this group suffers from this disease. However, even if no common features can be found, it is possible to obtain the privacy of people with disabilities if certain data exhibits tendency or similarity [55], which is known as ***Similarity at-***

tacks. In Figure 1.3, Mary's income is relatively low because no one in this group earns much. Another powerful inference attack is ***Composition attacks*** that utilize published datasets containing overlapping populations for cross-study to obtain personal information. It mainly targets partition-based anonymization methods, such as k-anonymity, but some randomization-based schemes, such as differential privacy, can provably resist this attack [56].

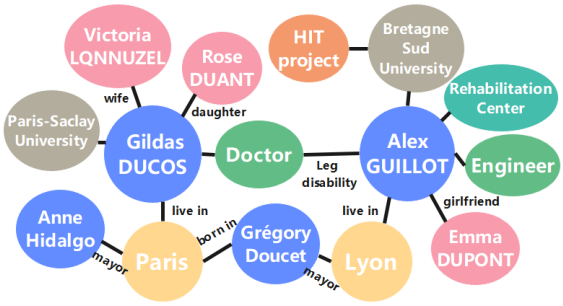


Figure 1.2 – Example of graph data



Figure 1.3 – Homogeneity attack and Similarity attack

Server-Ciphertext-Passive. Although the servers of healthcare institutions are generally more trustworthy and less likely to be threatened than those of ordinary services, the Cerballiance case [57] shows that its sensitive data can still be accessed without authorization. Once data is leaked, the consequences are even more severe, as healthcare organizations store large amounts of information about people with disabilities on their servers, including their identities and sensitive attributes. ***SCA*** could be used to steal their e-health information from the databases by, e.g., stealing the server's private key [58] and implementing covert channels [59]. Attackers may also exploit the cloud server's deduplication technology to infer whether a user has a disability or suffers from a certain disease [60]. In addition, traditional SCA use statistical tools for analysis. Today, with the advancement of ML and DL technologies, SCA based on these techniques are becoming popular [61]. The use of AI techniques has improved the accuracy of such attacks, broadened their applicability, and made them more threatening. Furthermore, most SCA are non-intrusive and passive, so they often leave no visible evidence after an attack, making such attacks difficult to prevent. Moreover, another attack ***TAO inference*** can compromise columns protected by property-revealing encryption in databases and further infer strongly encrypted or desensitized data through multi-column correlation and ML analysis. It has successfully inferred sensitive patient information from medical datasets,

e.g., personal information such as age and gender, and medical diagnoses such as mental health, substance abuse, chronic kidney disease, and leukemia [62].

Since the cloud stores a lot of sensitive data and has more computing resources (for analyzing the data, etc.), most of the attacks in this subsection are related to AI techniques, either targeting AI models or using AI to improve the accuracy of the attack. According to our research, even if the data are securely stored in the cloud, it should not be taken lightly because attacks here are more diverse and difficult to recognize.

1.1.3 Summary

In this section, we systematically analyzed a wide variety of attacks that can threaten the privacy of people with disabilities. From the above attack cases, it is clear that the vulnerability of e-health systems cannot be ignored. Therefore, it is especially important to develop effective defense measures. To facilitate a structured comparison of different attacks, Table 1.1 summarizes and analyzes their presence in the healthcare domain, as well as the types of data involved. Here, "literature presence" reflects the number of studies in the healthcare domain that address each type of attack, indicating how frequently the attack has been considered in this specific field. However, a low presence does not imply that the attack poses minimal risk; rather, it may indicate that the attack has not been sufficiently studied in the healthcare domain. This systematic organization and analysis provide a clear reference for the design of subsequent defense mechanisms and help to understand the relationships and differences among various attacks. Based on Table 1.1, inference attacks (including AI Privacy attacks and Inference attacks) are the most frequently observed or likely to occur in the healthcare domain, whereas other attacks, such as SCA, appear less frequently. However, this does not imply that SCA are impossible; due to their covert nature, these attacks may, in fact, pose a greater threat to the privacy of people with disabilities.

Table 1.1 – Comparative Overview of Potential Attacks in Healthcare.

Attack	**Literature Presence**	**Data Type**
Tunneling attacks	Low	Encoded textual data
Direct Access attacks	Low–Medium	Sensor data (numeric time-series)
DNS attacks	Low	Textual network data
SCA	Low–Medium	Various data (e.g., sensor, text, secret keys)
MITM attacks	Low	Plaintext / Encrypted network data
Replay attacks	Low	Encrypted network data
AI Privacy attacks	Medium–High	Health data (e.g., image), User attributes, Model parameters
Inference attacks	Medium	De-identified / Anonymized tabular and graph data

1.2 Privacy techniques

In the face of endless privacy attacks in e-health systems, proactive countermeasures are fundamental. This section focuses on various privacy protection techniques and strategies, including but not limited to, data encryption, anonymization, etc. To facilitate understanding and comparison of various privacy protection measures, our proposed taxonomy is shown in Figure 1.4. Our aim is to analyze their principles, advantages and limitations, and provide references for the security design and deployment of e-health systems. The state-of-the-art privacy-preserving mechanisms will be presented according to the above categorization. The following subsection is divided into three main parts: local, transmission, and server, and each part shows the defense countermeasures for plaintext and ciphertext. Before discussing them, it is necessary to introduce some basic terminology. For a given individual, structured personal data are categorized as follows [63]:

— *Explicit Identifiers (EIs)* directly identify individuals, e.g., social security number, ID number, name.

— *Quasi-Identifiers (QIs)* can be linked to externally published data to identify individuals, e.g., zip code, birthday, gender.

— *Sensitive Attributes (SAs)* are private data that the user does not want to be known, such as salary, disease history, and purchasing preference.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

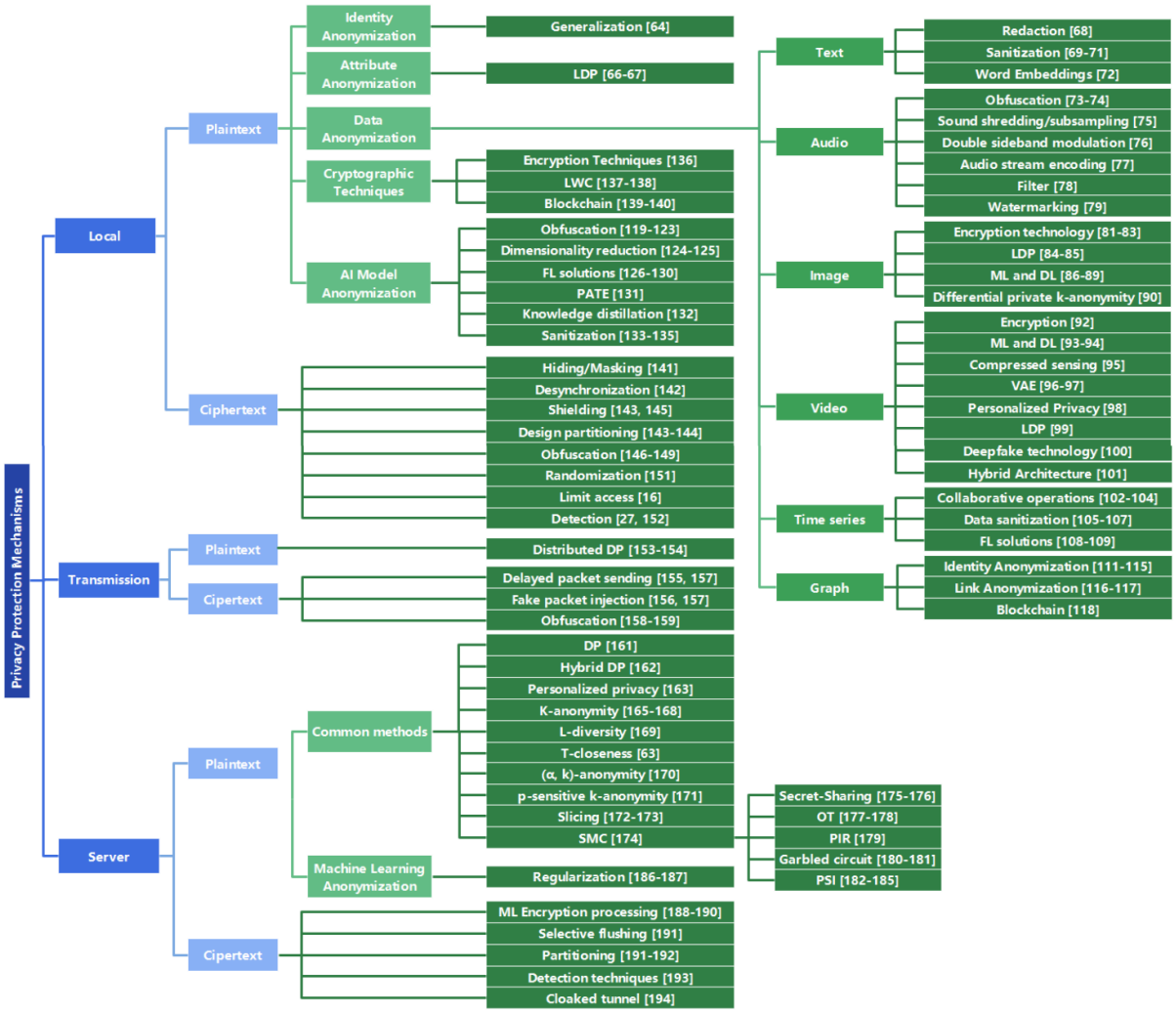


Figure 1.4 – Taxonomy of privacy protection mechanisms.

1.2.1 Local

Plaintext

This subsection analyzes various approaches to protect plaintext data. First, the focus will be on the privacy protection of EIs, QIs and SAs for structured data. Next, the protection mechanisms for unstructured data will be examined separately for different data types. Finally, encryption techniques and the privacy protection of AI models will be discussed.

[**Local-Plaintext**] **Identity Anonymization.** The identity protection algorithm is

presented here, and the data involved is the *EIs* mentioned above. Arguably, this type of information is the most serious to be compromised because attackers can learn about various privacy information such as disability status and other medical history based on the identity information. Even if all *EIs* are deleted before data is released, there is still a risk that they will be leaked. An effective measure against this threat is an identity anonymization approach combining k-anonymity with clustering [64]. The solution can be deployed in a local network of assisted living establishments, where smart devices from multiple residents (data owners) anonymize their health data in a distributed collaboration before sending them to a collector to ensure the identity privacy of data owners.

[Local-Plaintext] Attribute Anonymization. ***Local Differential Privacy*** *(LDP)* [65] adds random noise to user's local data, i.e., the deployment of privacy algorithm is entirely on the client side, thus enhancing the privacy of data in transit and in third-party servers. Healthcare-related LDP research [66] focuses on private statistical learning and inference, e.g., ML; statistical data analysis, e.g., data mining; and privacy amplification, which attempts to maximize data privacy without compromising data utility. LDP can protect attribute privacy well but not identity privacy, since attackers can use the server that collects the data to backtrack the input data to determine the specific data owner, while privacy-amplifying shuffling schemes prevent the data from being backtracked - the ***Encode, Shuffle, Analyze*** *(ESA)* architecture [67] guarantees data privacy through 3 steps: the encoder perturbs and encrypts the data, then the shuffler (serving as a trusted relay service) protects identity privacy by batch shuffling the data, and finally the analyzer prevents statistical inference attacks by analyzing the data.

[Local-Plaintext] Data Anonymization: Text. Text data is probably the most widely used data in our lives today, e.g., by people with disabilities who use chat apps to communicate with their doctors or entourage, or in diagnostic reports and prescriptions. Moreover, as an unstructured data, it is also more complex than structured data that have a clear classification. For textual data, it is necessary to first identify which texts lead to privacy breaches for the owner, and then determine the appropriate protection mechanisms based on the sensitivity of the texts. Another difficulty is that algorithms for training sensitive text categorization require providing predefined sensitive texts, sometimes some sensitive elements are ignored, while defining too many sensitive elements affects the usability of the text.

The act of removing or masking sensitive text is known in the literature as ***Redaction***. To this end, ERASE [68] allows sensitive words (e.g., diseases) in text to be dynamically

identified and removed, so that different users can see different content depending on their permissions. However, this approach affects the utility of the data [69]. In addition, the blacked-out portions can give attackers insight into the sensitivity of the text. Another method is ***Sanitization*** [69], which simply means automatically recognizing text-sensitive content (e.g., disability status, treatment), then protecting them by modifying/replacing them to hide private information of people with disabilities. A practical model for text sanitization is t-plausibility (t is a parameter representing the number of files) [70]. It desensitizes sensitive text by generalizing sensitive words to related words with broader semantics and keeps the document indistinguishable from other t-1 documents. However, it requires knowledge of the retrieved information to cover all sensitive entities, which limits the accuracy of sanitization and may generalize differently for different people, leading to differences in protection levels. Besides, it ignores the cases where semantic inference can leak sensitive information [71]. Thus, the C-sanitized model [71] is proposed, which is based on data semantics and information theory, and mimics the semantic inference that humans use to determine whether a text is sensitive and how to sanitize it. In addition, unlike t-plausibility, it does not require multiple documents and supports single-document cleanup. Differently, ***Word Embeddings*** [72] is language-independent and does not require manual annotation of training data. It identifies sensitive information about protected entities by learning detailed vector representations of language terms, uses vector similarity to measure semantic relevance, and replaces these sensitive texts in a generic way.

[Local-Plaintext] Data Anonymization: Audio. Audio data also contains private information that cannot be ignored. As an example, some people with disabilities wear IoT devices to monitor health status, e.g., respiratory rate, but these devices may also record persons' voice, such as conversations between them and family members, thus exposing their privacy.

For some wearable devices that provide services such as fall detection and cough detection, conversation is not as important as accurate detection of environmental sounds. Chen et al. [73] proposed a method to ***obfuscate*** speech intelligibility while preserving most of the environmental sounds. Based on the principle that purely vowel sentences are clearer than purely consonant sentences, it first recognizes the vocal tract regions of vowels and then replaces them with pre-recorded vowels, making words indistinguishable while preserving intonation and ambient sounds. Similarly, Larson et al. [74] ensured cough reconstruction and speech camouflage by transmitting the weights of some components,

but retaining the actual components needed to reconstruct the audio locally and on the server. Furthermore, ***Sound shredding*** destroys audio by extracting audio frames from the original sound frames and placing them randomly, whereas ***Sound subsampling*** captures sound frames randomly without preserving the full audio [75].

Recent research has focused on other aspects. [76, 77] have studied the privacy issues of recorded audio. Among them, [76] used ***double sideband modulation*** to modulate audible frequency waves into inaudible, non-recordable ultrasonic frequencies, thus protecting private conversations. However, sometimes there is a need to preserve the necessary medical information. In this case, Overo [77] records real-time audio streams in a standard AAC ***encoding*** format and balances privacy and utility according to the user's privacy requirements before releasing the audio, while ensuring that the recordings are not tampered with. In addition, there are cases where authorized apps record more information than allowed. Wang et al. [78] used a DNN algorithm to retain selected sensitive attributes (e.g., gender) but ***filter*** other information. Ultimately, the server can only access the obfuscated audio rather than raw data, thus protecting unauthorized data. Unlike obfuscation methods, Zhang et al. [79] proposed a reversible audio ***watermarking*** scheme to protect the privacy of telemedicine audio data. It decomposes medical audio into two separate embedding domains, embedding a robust watermark and a reversible watermark, respectively; and then utilizes the correlation between the sampling points to modify the histogram to reduce audio distortion and ensure high-quality audio data at the receiving end.

[Local-Plaintext] Data Anonymization: Image. In addition to personal information such as name, address, etc., image data may also reveal information such as portrait, disability status, etc., posing a greater threat to the privacy of persons with disabilities.

First, to ensure image privacy, ***encryption*** may be the first thing that comes to mind, but this poses certain difficulties for Content-Based Image Retrieval (CBIR). Xia et al. [80] provided a solution that extracts feature vectors to represent images (for retrieval) and protects them with the kNN algorithm. After the encrypted images are transferred to the cloud, a *watermark protocol* is used to prevent illegal copying and distribution of the retrieved images. Shen et al. [81] solved this problem with multiple image owners using *Secure Multi-party Computation (SMC)* techniques that allow each owner to encrypt the image features with a separate key. Furthermore, Yu et al. [82] utilized a multi-scroll memristive Hopfield neural network to generate highly complex chaotic encryption keys, combined with the MQTT protocol to enable secure transmission of medical data, ensuring

patient privacy protection during data sharing. Whereas MPVCNet [83] employs visual cryptography and trusted computing to achieve key-free encryption for medical images via distributed sharing and hardware-based security verification, and maintains AI recognition accuracy through transfer learning. In addition, ***LDP*** can also be applied to image data. To this end, the concept of m-neighborhood [84] is proposed, which protects sensitive information in adjacent images that differ by no more than m pixels. Shibata et al. [85] combined denoising diffusion probabilistic models and LDP to protect privacy in high-resolution medical images.

ML and ***DL*** are popular solutions for securely publishing image data. iPrivacy [86] is such a tool. It learns object-privacy correlations with the help of massive social media data to identify sensitive objects and blur them before sharing images. Similarly, Yu et al. [87] utilized *deep Convolutional Neural Network (CNN)* to detect privacy-sensitive content in images and combined *Generative Adversarial Networks (GANs)* and LDP to generate de-identified alternatives for privacy objects, while retaining image usability. To address the problem of being vulnerable to model inversion attacks in Federated Learning (FL) despite sharing only local model parameters, Li et al. [88] proposed using *selective parameter sharing* and LDP. There are also studies that focus on protecting the most private part of an image, i.e. the face. Among them, [89] relies on GANs to design a new generative model that can make the synthesized face blend well with the background, while increasing the difference between the synthesized face and the original face to prevent face re-identification. Cao et al. [90] proposed ***Differential Private k-anonymity***, which first finds the average face attributes of the face's k-nearest neighbors to hide face privacy while keeping small variations, then uses LDP to increase randomness and provide further protection. This scheme produces results that are closer to the original image and more natural than other GAN-based methods, such as DeepPrivacy or CIAGAN [90].

[Local-Plaintext] Data Anonymization: Video. Video data is common in healthcare, especially for people with disabilities. If they have an accident at home, it may be too late to go to the hospital, or the injuries may not be serious enough, but their family or entourage does not know what to do. The fastest and most effective solution is to call a doctor via video. However, people still hope to receive support in their daily lives without compromising their privacy.

Some traditional methods, e.g., blurring and pixelating, have a high risk of face recognition even after application [91]; whereas scrambling and morphing are more effective but still have the threat of reconstruction attacks, so some schemes combine them with

encryption [92]. In recent years, researchers have applied ***AI*** techniques to video privacy protection. Ren et al. [93] extended GANs with multiple tasks. They designed a face anonymizer to hide face information in videos but maximize spatial action information, and a face recognizer to extract sensitive information from anonymized videos. PrivacyPhys [94] uses a pre-trained 3D CNN to modify invisible remote photoplethysmography in facial videos, thus preventing disease detection, emotion recognition, and biometric identification. Another approach to protect privacy in video surveillance is ***compressed sensing***, which efficiently compresses video data and recognizes action information in compressed videos using DL algorithms. For instance, [95] combines CNN and RNN to obtain spatio-temporal features of compressed videos. Several studies focus on extensions of the ***Variational Autoencoder*** *(VAE)*. Xu et al. [96] proposed a cycle vector-quantized VAE (cycle-VQ-VAE) framework for encoding and decoding video and its extracted audio, and combined it with video compression techniques to prevent SCA during video transmission. VQ-VAE is also used in the Quality Maintenance-VAE proposal by Qiu et al. [97]. It automatically adjusts the generated images to different quality raters.

In addition to face privacy, privacy leakage due to video context cannot be ignored. Du et al. [98] proposed a concept map to characterize individual privacy needs and implement ***personalized privacy*** policies based on different needs. ***LDP*** [99] can also be applied to video data to quantify and limit privacy leakage, and these schemes consider sensitive objects other than people. Furthermore, another study [100] notes scenarios where facial information needs to be preserved, such as facial injuries of people with disabilities or telemedicine diagnosis. For this reason, ***deepfake technology*** is proposed, which outperforms traditional methods because facial keypoints remain almost unchanged after face-swapping. In addition, deployment of complex algorithms can be difficult in real-life situations, especially in edge computing, e.g., on edge cameras used to monitor people with disabilities. Hence, a ***hybrid architecture***, PriSE [101], is proposed that uses a lightweight foreground object scanner that scrambles frames containing foreground objects and detects faces using a multitask cascading CNN.

[Local-Plaintext] Data Anonymization: Time series. Smart wearable devices are often used to measure time-series data such as body temperature, heart rate, blood pressure, etc. in real time to monitor the physical condition of people with disabilities. The threat to their privacy posed by these data is often overlooked, but the fact remains that malicious adversaries can still use these data to analyze their private information such as gender, age, and disability status. To protect the privacy of these data, the following

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

discussion will outline several mainstream methods, including Collaborative operations, Data sanitization and Federated learning. It should be clarified that these methods are applied to raw sensor data rather than extracted features.

For ***collaborative operations***, Shi et al. [102] proposed a novel private stream aggregation method that combines HE and LDP. Users periodically upload encrypted, noisy data to an untrusted data aggregator that is able to compute over multiple cycles, so it can decipher the noisy sum of all individuals but cannot infer extra information about each individual. Joye et al. [103] optimized the method to support large plaintext spaces and a large number of users, improving the efficiency of the algorithm and allowing it to be used in both offline and online modes. LDP is implemented based on adding random noise, but for physiological time series such as respiration and ECG, small perturbations may affect the data utility. Additionally, Lyu et al. [104] proposed the RG-RP scheme and combined it with a novel LSTM-CNN model to realize a privacy-preserving collaborative learning framework. The RG-RP scheme is divided into two stages: in the first stage, it utilizes a nonlinear function to perturb the data; in the second stage, it compresses the data by a row-orthogonal random projection matrix. This framework achieves a favorable balance between privacy protection and recognition accuracy. Another approach, ***data sanitization***, is based on the core idea of replacing sensitive information with non-sensitive information. Malekzadeh et al. [105] developed Replacement Autoencoder (RAE) for disguising sensitive information (i.e., gender) in time-series data. The RAE introduces a deep autoencoder architecture to first extract discriminative features of data corresponding to sensitive inferences from time-series data. The algorithm then learns how to replace these sensitive features with features commonly found in non-sensitive inferences and preserves the information needed for activity recognition in the data. As a result, the RAE maintains recognition accuracy while protecting the privacy of sensitive information. Based on this, they extended their work and introduced the Guardian-Estimator-Neutralizer framework for sharing only transformed versions of the sensor data [106]. The Guardian is a feature learning framework that uses a deep autoencoder to identify and differentiate discerning features within the data. The Estimator utilizes a multitask convolutional neural network to evaluate the accuracy of an algorithm in making both sensitive and non-sensitive inferences on the transformed data. The Neutralizer is an optimizer that assists the Guardian in finding the optimal transformation function to infer activities without revealing sensitive information. Building on their works, Bigelli et al. [107] developed a multiple attributes obfuscator that protects gender, age, height and weight information

simultaneously. In addition, regarding ***federated learning*** solutions, HARFLS [108] uses FederatedAveraging method to share the model weights and HE is applied to avoid the leakage of weight information. Whereas [109] focuses on the federated unlearning mechanism, enabling the model to effectively remove the influence of withdrawn data when users request it, thereby enhancing data controllability.

[Local-Plaintext] Data Anonymization: Graph. Graph data, see Figure 1.2, contains very rich information and is therefore widely used for personalized recommendations, e.g., content or services. Personalized healthcare services help people with disabilities manage their daily health and provide appropriate rehabilitation plans based on their conditions. However, the data may contain private information that they do not wish to disclose, e.g., disability status. In addition, PHRs represented as graph data may lead to the disclosure of medical knowledge [110], which can threaten not only the privacy of people with disabilities, but also that of physicians and their interactions.

The first step is ***Identity anonymization***. *K-degree anonymous graph* [111] protects identity privacy by adding or removing a minimum number of edges so that each vertex in the graph has the same degree as at least k-1 other vertices. *K-automorphism* [112] is an extended version of this, and Skarkala et al. [113] considered weighted graphs, i.e., graphs in which link strength matters. Other ideas include anonymizing individuals into clusters or classes [114], *DP* [115], etc. However, even if identities are effectively hidden, attackers can still infer sensitive relationships between individuals, so ***Link anonymization*** is also needed. Zheleva et al. [116] proposed five different privacy-preserving approaches to tune the data deletion and privacy-preserving ratios affecting data utility. Mittal et al. [117] proposed a *random walk based* link anonymization technique with secure routing. Furthermore, Elhoseny et al. [118] used ***Blockchain*** to protect the privacy of medical data based on graph data modeling in healthcare scenarios. The solution optimizes data transfer paths through intelligent graph data modeling and combines digital certificate-based authentication and blockchain technology to achieve privacy and integrity of big medical data.

[Local-Plaintext] AI model Anonymization. To defend against AI model inference attacks, LDP is an ***obfuscation*** method that can effectively protect the privacy of e-health data in AI models. LDP-FL allows for training models in a distributed environment while protecting local data privacy. With enough participants, Geyer et al. [119] used LDP to hide the user contribution and information during data training, while ensuring model performance. LATENT [120] enhances e-health data privacy in the IoT environ-

ment by adding a randomization layer between the convolutional and fully-connected layers of a CNN. EdgeSanitizer [121] is an edge computing approach based on a DL inference framework for resource-constrained edge devices that adaptively injects noise to blur learned features to protect private information in mobile data analytics. To address the need for training larger scale AI models using highly sensitive data (such as hospital-owned local servers or private clouds), Arasteh et al. [122] introduced DP into the model training process to build AI models that balance privacy, accuracy and fairness to fully protect the privacy of individual's sensitive medical records, as well as patient identity and sensitive attributes. Additionally, considering collaborative medical AI scenarios, Tayebi et al. [123] suggested adding calibrated noise to gradients during model training to address the challenge of FL being vulnerable to membership inference attacks (MIA) and reconstruction attacks. Another method is ***Dimensionality reduction***, which perturbs the data by projecting it to a low-dimensional hyperplane, i.e., privacy is preserved by discarding some information. Random matrices [124] can be used to reduce the dimensionality of input data, but the results are somewhat similar to the original data. Based on this, Kung et al. [125] proposed to project data in a way that maximizes practical label accuracy while reducing private label accuracy, which prevents data leakage to some extent.

In assisted living scenarios such as health monitoring, smart devices often collaboratively build models through distributed learning, e.g., ***FL***, to provide personalized services without sharing raw data. Despite enhanced privacy protection, this collaborative process is still vulnerable to threats such as reconstruction attacks. Therefore, FL is usually combined with ***DP***. For example, Shokri et al. [126] proposed to selectively share model parameters and apply DP in parameter updates to address privacy threats in sensitive data collection. Besides, the combination of FL and ***Cryptographic techniques*** [127] is also widely adopted, especially in the medical field. Furthermore, Kumar et al. [128] proposed a FL framework based on ***Blockchain*** and ***HE***, which aims to protect the privacy of medical image data in healthcare scenarios. It enables decentralized and secure model aggregation and training, avoiding the direct sharing of sensitive data. Moreover, ***K-anonymity*** can also be combined with FL to anonymize user identity and protect health data and location privacy in smart healthcare systems [129]. A more recent comprehensive scheme [130] uses an integrated framework that combines FL, SMC, HE, and DP to protect data privacy in AI healthcare systems through local data encryption, distributed model training, and dynamic encryption tuning.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

In addition, to defend against MIA, two types of distributed privacy-preserving learning methods can be employed: ***Private Aggregation of Teacher Ensembles*** *(PATE)* [131] achieves strict DP by allowing multiple 'teacher' models trained on a subset of sensitive data to instruct a 'student' model that does not have direct access to sensitive data in a noisy voting manner. While the distillation for membership privacy [132] uses ***Knowledge distillation*** to transfer knowledge from sensitive data-trained models to protected models, thus avoiding direct access to the original private data. In this way, it protects sensitive data, such as medical records, against MIA. Regarding secret extraction attacks, i.e., attacks on large amounts of text in AI models, methods such as ***Sanitization*** [133] applied to text data can also be effective against these attacks. Finally, AI healthcare privacy concerns and privacy-preserving techniques have been thoroughly reviewed in [134, 135].

[**Local-Plaintext**] **Cryptographic Techniques.** As mentioned above, encryption or cryptography is very powerful in solving privacy issues. Among them, ***HE*** [136] is popular because it allows operations to be performed directly on the ciphertext with the same result as if the operations were performed on the plaintext and then encrypted. However, given the resource constraints of IoT devices, traditional cryptographic protocols may not be applicable to all IoT environments even if the encryption technology is very strong, so ***Lightweight cryptography*** *(LWC)* [137] has emerged. It can meet security and privacy requirements in resource-constrained scenarios, such as smart healthcare [138]. Although LWC offers efficiency advantages, it typically does not support direct processing on encrypted data. As a result, decryption is required before processing, which poses certain limitations on privacy protection. In addition, cryptography-based ***Blockchain*** is commonly used in smart healthcare, because its secure, decentralized, and transparent nature allows authorized parties to jointly access and manage the health data of people with disabilities, and the records cannot be altered or deleted. Based on blockchain technology, researchers have devised many schemes for managing healthcare-related private data that strike a balance between data privacy and data accessibility, which are well summarized in the following literature [139, 140].

Ciphertext

This section discusses various defenses against local Side-Channel Attacks (SCA).

[**Local-Ciphertext**] First, to defense against power analysis attacks, hiding or masking strategies can be used. ***Hiding*** schemes typically change the power consumption for

each operation by using dual-rail precharge, asynchronous and current-mode logic styles to keep it approximately the same [141]. ***Masking*** schemes randomize the power consumption of the encryption process by using masked dual-rail precharge logic, random switching logic, adding noise or filtering circuitry [141]. Another countermeasure is to ***desynchronize*** the power traces using false cycle insertion, unstable clocking, or random delays [142] so that they are no longer aligned in the same collection set. A straightforward way to defend against electromagnetic (EM) analysis is ***shielding*** EM radiation to reduce leakage, or ***reducing the overall power consumption*** of the chip to suppress the emission of EM signals [143]. Since EM signals containing private data may be coupled to other nodes and lead to information leakage, ***design partitioning*** is proposed to separate the plaintext and ciphertext zones on the chip. In addition to this, typical shared infrastructures such as power or clock can be separated [143]. The ***3D die stacking*** technology allows for a more thorough separation of the two zones, which are actually deployed in different chips that are then glued and stacked together [144]. Furthermore, ***shielding*** solutions are also effective for acoustic analysis and can be achieved by using anechoic chamber materials in the enclosure of the equipment [145]. Another solution is to ***mix random noises*** of similar frequency or background sounds with acoustic emission, making it difficult to identify usable information [146]. Additionally, AudioSentry [147] utilizes existing equipment to create ***distributed obfuscator arrays*** that generate obfuscation beams in different directions, thereby effectively stopping multi-microphone adversaries from stealing user privacy while still retaining authorized perceptual capabilities.

Finally, we discuss how to defend against smartphone-based SCA. First, ***random vibrators*** [148] can be used to prevent keystroke inference attacks, which can infer typed inputs, such as diseases, from covert video recordings of backside motion. For sensor-based attacks, this scheme cannot provide protection, whereas ***injecting noise*** into sensor readings [149] offers effective protection. A more general approach is to inject noise into the information provided via the proc file system [150]. Second, ***keyboard layout randomization*** [151] is also a countermeasure that can prevent attackers from inferring keyboard inputs using sensor readings. It has been suggested to ***limit sensor access or sampling frequency*** of sensitive inputs, but this may limit the functionality of some apps [16]. Some researchers have designed applications [152] or frameworks [27] that can ***detect*** different SCA on mobile devices. Among them, [152] can also recognize pure software attacks.

1.2.2 Transmission/Transport

Data transmission is an integral part of e-Health IoT systems, and traditional data transmission protection schemes are becoming increasingly vulnerable in the face of evolving cyber threats, coupled with the importance of health-related data, all of which compel us to focus on innovative and more robust security enhancement approaches in this area to protect the privacy and security of health data during network transmission.

Plaintext

Distributed differential privacy *(DDP)* first aggregates the data sent by the client at multiple trusted intermediate nodes, then encrypts or disturbs this data and transmits it to the server, which ultimately can only have access to the aggregation results, but not the original data [153]. In addition, it is often combined with methods such as SMC [154] to further improve the protection of the privacy of people with disabilities. In this way, both data privacy and identity privacy of people with disabilities are protected. In fact, the privacy protection of the plaintext in the data transmission phase is mostly accomplished with the help of locally deployed DP or encryption solutions, etc.

Ciphertext

To defend against timing-based network SCA, such as behavioral inference, fingerprinting, and timing-based snooping attacks, protection strategies fall into two categories: ***delayed packet sending*** and ***fake packet injection***, which break the temporal correlation between events and signals. Delay schemes use a random interval generator to delay packet transmissions for a random amount of time, but they are less suitable for delay-sensitive healthcare surveillance systems [155]. Injection schemes protect real packets by randomly injecting fake packets with the same characteristics as real packets into network traffic, making it difficult for attackers to distinguish them [156]. There are also hybrid solutions that combine the two to enhance privacy protection, and [157] also considers the impact of response times of different IoT devices. In addition to timing analysis, for traffic analysis based on packet information, ***obfuscation*** is often used to hide this information. An approach is to *uniform packet size* over independent link padding [158]. Another option is *morphing*, which blocks attacks by modifying one packet in the traffic to make it look like another [159]. Xiong et al. [160] used a packet obfuscator with discrete memoryless channels for padding to make the size set of obfuscated packets sent by IoT

Table 1.2 – Comparison of CDP, LDP, DDP and HDP

Category	Advantages	Disadvantages	Action Position
CDP	High accuracy; easy to implement	Requires trusted server; risk of data exposure	Server (centralized)
LDP	Strong privacy; no trust in server	Low accuracy; high noise	Local (user device)
DDP	No central trust; better accuracy than LDP	Complex communication / computation; weak trust assumptions for some implementations	Multiple distributed entities
HDP	Flexibility; privacy-accuracy trade-off	Complex to implement	Local + Server

devices indistinguishable from the original packets.

1.2.3 Server/Cloud

This section discusses measures to protect sensitive information of people with disabilities on the server. These methods can also be roughly divided into plaintext processing and ciphertext processing.

Plaintext

Differential privacy *(DP)* [161] was first proposed in a centralized form, where raw data are sent to a trusted data center and then scrambled (usually by adding random noise), so that the results of queries on two datasets with only one distinct record would be highly similar, and attackers cannot learn about changes to a single record in the dataset from the changes in the output. As a result, the privacy of people with disabilities is protected. Furthermore, Avent et al. [162] proposed a ***Hybrid DP*** *(HDP)*. It categorizes users according to their trust in the server: users with the lowest trust use LDP with the lowest privacy budget, while users with the highest trust can send raw data, and the server processes the data based on the trust. Therefore, additional preprocessing is required to classify the trust relationship between users and servers. Finally, Table 1.2 summarizes the advantages and disadvantages of centralized DP (CDP), LDP, DDP and HDP.

In real life, everyone has different needs for privacy protection and may have different sensitivities to the same sensitive information, thus giving rise to the concept of ***Personalized privacy***. Simply put, it defines different privacy levels, e.g., low, medium, high, so

that people with disabilities can customize the privacy level of their data, then the algorithm applies different privacy policies for different levels to better protect privacy in the release of high-dimensional healthcare data [163]. In addition, CIPPPA [164] protects identity privacy through anonymous identity generation and condition tracking mechanisms, ensuring that patient identities are not visible to cloud servers and auditors. Next, the purpose of ***k-anonymity*** is to anonymize data before it is published. As an example, for data such as physiological signals from wearable IoT devices, Guo et al. [165] significantly reduced the potential privacy exposure of people with disabilities by applying personalized k-anonymity for static and long-term data, and temporal DP for real-time data. Moreover, k-anonymity can anonymize various data streams from IoT devices in conjunction with *partitioning* techniques [166] and can serve authentication scenarios in the cloud [167]. The biggest challenge on this topic is the choice of k, i.e., the need to balance data privacy and data utility, for which Arava et al. [168] provided interesting practices. However, as shown in Figure 1.3, k-anonymity prevents identity disclosure but not attribute disclosure because it sometimes fails to resist homogeneity attacks. Therefore, more powerful solutions have been proposed based on satisfying k-anonymity. ***l-diversity*** [169] requires that sensitive attributes in each equivalence class in a dataset have at least l distinguishable values. However, when a disease has only two attributes, "positive" and "negative", and only 1% of the data is positive, it is vulnerable to similarity attacks. ***T-closeness*** [63] solves this problem by ensuring that the distribution of sensitive attributes in each equivalence class of the dataset is close to the distribution of sensitive attributes in the whole dataset and their distance does not exceed a threshold t, thus protecting these sensitive attributes. Another improvement is ***(α, k)-anonymity*** [170], which controls the probability of occurrence of each sensitive attribute in each equivalence class to be not greater than α. While ***p-sensitive k-anonymity*** [171] requires that there are at least p different sensitive attributes in each equivalence class with $p \leq k$. However, the higher the data dimension, the harder these schemes are to implement.

In contrast, ***slicing*** (for data publishing) [172] protects privacy by partitioning data horizontally (grouping data records) and vertically (combining highly correlated attributes, such as diseases and treatments) in both directions. Compared to the above methods, *slicing* better preserves data utility, effectively prevents the disclosure of sensitive attributes and memberships, and is suitable for handling high-dimensional data. However, considering the unreliability of random permutations, an enhancement approach [173] combining upper and lower protection levels and value swapping has been proposed to better handle

unique and highly identical attribute values in e-health data releases to further reduce the risk of composition attacks.

[Server-Plaintext] Cryptographic Techniques. Encryption is very powerful, however, in addition to data confidentiality, data utility must also be considered, because doctors need the necessary information for medical diagnosis. ***Secure Multi-party Computation*** *(SMC)* [174] allows sharing of "secrets" in semi- or low-trusted environments. Multiple parties have access to the correct computational results, but not to any information beyond the results, thus ensuring the security and privacy of the parties' e-health data. ***Secret-sharing*** [175] is a fundamental technology in SMC. It can split sensitive data (e.g., health data) into multiple shares that are stored in different organizations, and only a union of participants that meets a threshold can reconstruct the original data. Therefore, it is mainly used to securely store private data, e.g., personal health records, in the cloud. The scheme [176] divides sensitive data to be stored in different cloud providers and then outsources the complex data reconstruction to cloud computing service providers. Even if a particular cloud storage fails or is attacked, the security of the data is guaranteed. ***Oblivious Transfer*** *(OT)* [177] is another fundamental technology of SMC, which allows the data receiver to obtain part of the data from the data provider without revealing the content of the selection, and at the same time the data provider will not know which part the receiver has obtained. The k-out-of-n OT is usually used due to the complexity of real-world scenarios [178]. It enables secure queries, e.g., access to medical records without revealing the patient's specific needs.

Although OT ensures mutual privacy in small-scale exchanges, ***Privacy Information Retrieval*** *(PIR)* [179] extends similar protections to large-scale data queries in healthcare databases. In addition, PIR is more focused on receiver privacy. As an OT-based encryption method, ***Garbled circuit*** ensures that multiple participants jointly compute medical models or diagnostic results without disclosing their private data. For example, Chen et al. [180] proposed to compare the statistics of different healthcare organizations to improve their service quality. In smart healthcare, to ensure the privacy of sensitive physiological data (e.g., ECG) during the quality assessment process, Barni et al. [181] designed a hybrid multi-party computation protocol based on HE and Yao's garbled circuits. Another similar concept is ***Private Set Intersection*** *(PSI)* [182], which allows both parties holding data to compute the intersection of their respective datasets without revealing any information other than the intersection. In healthcare, it can provide services while protecting the privacy of people with disabilities, e.g., archiving medical

information of people with similar disabilities [183], sharing PHRs from different healthcare providers with doctors [184]. Moreover, given the complexity of real-life scenarios, multi-party PSI has been studied [185].

[Server-Plaintext] AI models Anonymization. To defend against Membership Inference Attacks (MIA) of AI models, the straightforward confidence masking method hides the true confidence score by providing only partial prediction information or adding noise to the prediction vector, but it cannot adequately defend against MIA [45]. Another method, ***Regularization***, can be effective in providing protection that defends against MIA by mitigating model overfitting to protect patient privacy. Adversarial regularization [186] makes the model's predictions on training and non-training data indistinguishable by formalizing the problem as a min-max game, while mixup+MMD [187] combines a maximum mean discrepancy (MMD) regularizer and mix-up training with the aim of resisting attacks by reducing the model's generalization gap.

Ciphertext

By applying local encryption to sensitive data, e.g., HE, data security is guaranteed, but data utility is greatly reduced. For this reason, Bost et al. [188] proposed an efficient privacy-preserving protocol for ML classification on encrypted data in sensitive scenarios such as medical prediction, which ensures user data and model confidentiality by building modules to perform classification directly on homomorphic encrypted data. Whereas CryptoNets [189] focuses on neural networks, which can convert trained neural networks into networks that can operate efficiently directly on encrypted data, thus enabling ***Privacy-preserving prediction*** in the cloud. Another work [190] proposed to process and predict encrypted human skeletal data directly in the cloud without decryption through full HE. It is mainly applied to remote medical monitoring, e.g., fall detection.

The solution for server Side-Channel Attacks (SCA) is different from local attacks, but the general idea is similar. To defend against sequential cache SCA, Godfrey et al. [191] proposed a ***selective cache flushing*** scheme that flushes the cache only when a context switches from one domain to another, thus preventing attackers from accessing private cache data. For parallel cache SCA, the ***cache partitioning*** [191, 192] schemes are used. These schemes defend against attacks by giving virtual machines (VMs) exclusive access to portions of their caches (called secure or stealth pages), thus preventing attackers from inferring which sensitive EHRs are being accessed by healthcare workers through

the shared cache. CloudRadar [193] uses signature- and anomaly-based ***detection techniques*** to identify cache-based SCA. Whereas Pacer [194] eliminates side-channel leaks through a novel ***cloaked tunnel*** abstraction that shapes client traffic externally to the VM, making its timing and size characteristics independent of secret information, thereby providing end-to-end privacy protection.

1.2.4 Summary

In general, DP and encryption-based schemes have become the most widely adopted privacy-preserving techniques due to their flexibility and robustness. In addition, blockchain and SMC provide reliable frameworks for ensuring data integrity and privacy in distributed systems. Consequently, FL enhanced with DP, encryption, blockchain, or SMC has emerged as a popular solution that offers more comprehensive privacy protection for people with disabilities.

However, these approaches still face notable limitations, particularly in IoT environments. Encryption-based schemes often incur substantial computational and communication overhead, making them impractical for resource-constrained edge devices, such as smartwatches, fitness trackers, and medical sensors. Similarly, FL frameworks typically assume stable connectivity and sufficient local resources, conditions that may not hold in many real-world IoT deployments.

Other lightweight techniques, such as data sanitization or masking, are more suitable for local deployment but provide only limited privacy protection. These methods generally replace predefined sensitive attributes or keywords while transmitting the remaining data in its original form. Consequently, unrecognized or context-dependent sensitive information may still be exposed, leading to potential privacy leakage.

A further limitation lies in the trade-off between privacy protection and data utility. Methods such as DP or noise-based obfuscation add random noise to protect sensitive information, but this process inevitably reduces the accuracy and reliability of the collected data (especially for time-series data). In IoT applications where tasks like health monitoring or activity recognition require precise and continuous signals, such loss of data quality can significantly affect model performance and decision-making. Therefore, these methods are often unsuitable for edge devices that depend on high-fidelity data and real-time processing. In the next section, the challenges and research directions that this thesis focuses on will be analyzed and discussed.

1.3 Conclusion

This chapter comprehensively summarizes the various privacy threats faced by people with disabilities in assisted living scenarios and their corresponding protection countermeasures, and discusses the various types of data that may be involved in this scenario, including but not limited to text, audio, images, etc. Based on this, we have identified some limitations in current solutions, and these shortcomings are precisely the focus of our research. This thesis aims to address some of these limitations, which are described in the following. In addition, it is important to emphasize that our focus is on privacy protection issues related to time series data generated by sensor measurements.

1.3.1 Local deployment constraints in real-world scenarios

With the popularity of assisted living applications, effective privacy protection for users (especially special groups, e.g., people with disabilities) has become a central concern. However, current practices face a serious challenge: many privacy protection schemes, such as encryption or FL combined with DP, have difficulty running efficiently on resource-constrained local environments (i.e., IoT devices and edge servers) due to their high computation and storage overheads. This challenge is even more significant for heterogeneous IoT devices, such as commonly used wearables for health monitoring like smartwatches and fitness trackers, which differ significantly in computing power, storage capacity, and energy consumption. In monitoring scenarios that require high real-time performance, these limitations make it difficult for devices to simultaneously meet performance and privacy protection requirements. In addition, the risk of privacy leakage is ubiquitous throughout the entire process of data collection, transmission, storage and application. Therefore, deploying privacy protection mechanisms locally on IoT devices is an inevitable choice to solve the above dilemma. Local deployment not only fundamentally improves the level of privacy protection, but is also the key to building trust in assisted living scenarios and driving the popularization of the technology. To this end, this thesis aims to design a lightweight, locally deployable privacy-preserving solution that can accommodate various IoT devices while ensuring secure, efficient, and reliable data processing.

1.3.2 Robust privacy-preserving solutions

In the complex environment of the real world, user privacy is threatened by a variety of potential attackers. These attackers may include malicious insiders, external hackers, and entities that may gain access through improper data sharing. According to our survey and analysis of existing research, many current privacy-preserving schemes fail to fully account for the diversity and interactions of possible attacker models, leaving potential risks of privacy leakage when exposed to complex threats. To address this limitation, this thesis will systematically consider a wide range of possible attack scenarios and, based on these considerations, design a more robust privacy-preserving solution. This approach integrates multiple layers of defense from the design stage, taking into account different attacker capabilities and objectives, thereby enabling more comprehensive and sustainable privacy protection.

1.3.3 Lack of datasets that include people with disabilities

The ultimate goal of assisted living technology is to enhance the quality of life for all individuals, however, a significant shortcoming of existing human activity recognition datasets is that they fail to adequately reflect the activity patterns and specific needs of people with disabilities. This data homogenization severely limits the development of assisted living technologies and makes it difficult for them to truly benefit all types of users. In order to advance technology in this field and ensure its inclusiveness and effectiveness, there is an urgent need to significantly increase the breadth and depth of research datasets. This means that future datasets must encompass a more diverse group of participants, including people with all types of disabilities, and be designed with rich and varied scenarios of daily activities. At the same time, the amount of data and the number of participants need to reach a sufficient scale to support effective training and generalization of the algorithms. More importantly, these diverse and high-quality datasets will provide a valuable foundation for designing and validating more robust privacy-preserving solutions, ensuring that assisted living technologies serve the needs of all people while protecting user privacy.

Chapter 2

Attacker Models and the Datasets Used

After exploring various state-of-the-art privacy attacks and defense mechanisms in assisted living, this chapter will focus on the attacker model considered in this thesis. Furthermore, it should be particularly emphasized that all the work discussed in this thesis is situated within the Human Activity Recognition (HAR) framework and the data involved is time-series data. This point will be consistent throughout the subsequent discussion. The main goal of the attacker in our research scenario is to train a Machine Learning / Deep Learning (ML / DL) algorithm that is capable of recognizing sensitive information of individuals with disabilities based on time-series data obtained from the cloud. After the data is transmitted to the cloud, it undergoes multiple processing procedures, and an attacker may exploit some of them for the purpose of stealing data. The data processing procedures in the cloud and the attacker model are as shown in Figure 2.1.

The ultimate goal of the attacker is to be able to infer sensitive information from any data sample, which corresponds to the server-side identity inference attack and attribute inference attack in our taxonomy. Specifically, the attacker is assumed to be an external entity with powerful computational resources and expertise that does not directly control the data source, but is able to obtain time-series data related to individuals with disabilities that are stored in the cloud through eavesdropping or legitimate but malicious means of access, and potentially even directly obtain the labels corresponding to the data. They have the ability to construct complex ML/DL models designed to identify individuals from these data (i.e., identity inference attack) or infer sensitive attributes of the individual, such as the type of their disability, severity, etc. (i.e., attribute inference attack). In addition, it is important to clarify that this thesis focuses primarily on identity inference attacks. However, in our experiments, and often in real-world scenarios, the user identity labels associated with the data are represented by user IDs or pseudonyms rather than real

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

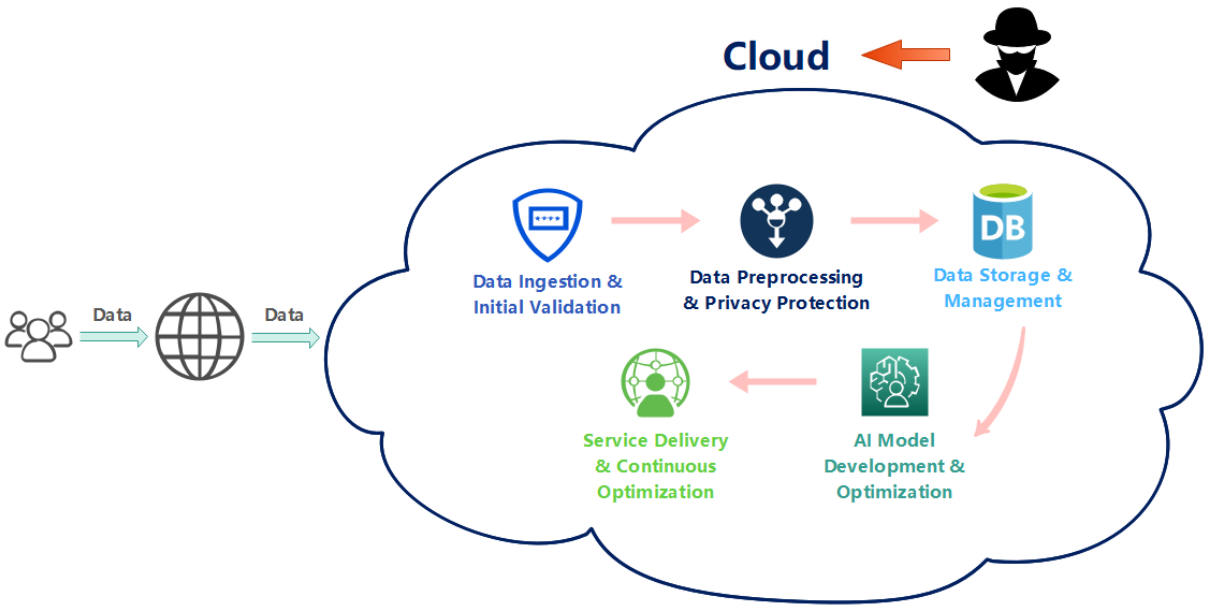


Figure 2.1 – Data Processing Stages in Home Healthcare Systems.

names. However, when an attacker compromises an anonymized database that still contains such identity-related labels, these labels may no longer remain protected, potentially allowing the attacker to link them with real user identities. In this case, once the attacker gains access to both the data and their corresponding labels, they can train powerful Artificial Intelligence (AI) inference models capable of inferring sensitive information from unseen or newly collected data, and possibly even deducing additional personal attributes beyond identity from the intercepted data. Attackers could use this inferred information for targeted marketing, social discrimination, or even more malicious purposes. Hence, the defense strategy in this thesis will focus on how to protect the privacy of time-series data of individuals with disabilities under such attacks.

Therefore, Section 2.1 will present the attacker models considered in this thesis, and the various potential attacks can be categorized as follows, depending on the amount of information available to the different attackers: supervised learning attacks in Section 2.1.1, semi-supervised learning attacks in Section 2.1.2, and unsupervised learning attacks in Section 2.1.3. Next, the datasets used to validate our algorithms, i.e., HASC-IPSC, WISDM, will be described in Section 2.2.1 and 2.2.2, respectively.

2.1 Attacker Models

2.1.1 Supervised Learning Attacks

Principles of Attack

In HAR systems, the threat posed reaches its highest point when an attacker is able to simultaneously access all (or most) users' sensitive sensor data (e.g., acceleration, gyroscope, physiological signal readings such as heart rates, etc., from smart wearable or mobile devices) as well as all (or most) of the corresponding sensitive attribute labels. These sensitive attributes are primarily personal information, such as user identifiers, age, gender, height, and even more private health conditions (e.g., disability status, etc.) that are closely associated with these behavioral data. In this ideal but dangerous scenario, attackers have all the elements required to train supervised learning models: for example, by directly obtaining access to a sensitive database containing samples and labels; or by acquiring an external database with a high user overlap with the target database, and they can then infer target users' identities via record linkage and statistical matching (pairing records across databases), feature co-occurrence comparison (comparing rare feature combinations), etc. Subsequently, attackers can use these time-series data with real labels to train highly accurate ML or DL models. Once constructed, these models are able to attribute new, unlabeled sensor data and allow attackers not only to accurately identify and track users, but also to deeply analyze and predict their behavioral patterns, life routines, and so on, thus posing a serious and far-reaching threat to the privacy of people with disabilities. In this extreme case of "total data transparency", the security of people with disabilities' data would be completely dismantled and the foundation of trust in the HAR system would collapse. The possible attack scenarios and specific examples are shown in the Figure 2.2 and will be described in detail in the next subsection.

Attack Scenarios and Examples

There are various ways for attackers to penetrate HAR systems to gain access to the sensitive data, as shown in Figure 2.2. One of the main ways is a direct attack against the core data storage infrastructure. This is typically manifested as a **compromise of the core database of the cloud service provider**. This database is the central repository that stores sensor data, precise activity labels, identity information, etc. of users. Attackers may be able to batch export the complete HAR data of all users by exploiting known

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

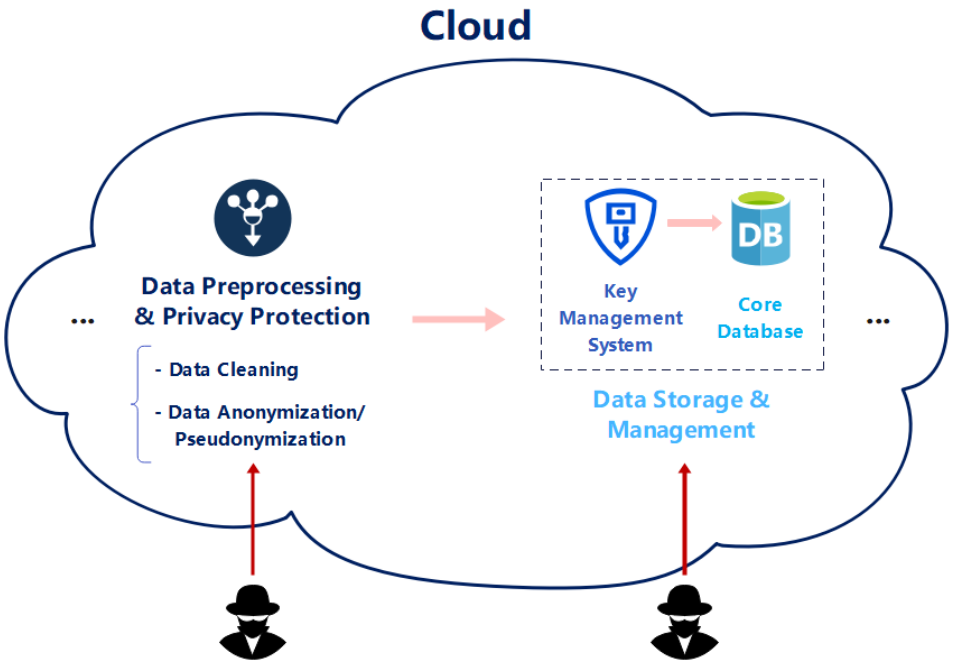


Figure 2.2 – Examples of Supervised Learning Attacks in the Cloud.

software vulnerabilities (e.g., unpatched vulnerabilities found in database management systems or related applications), executing *malicious injection attacks* (e.g., SQL injections, which bypass security restrictions by submitting specially crafted data queries), or gaining direct access to the database using stolen or weakly processed *administrator credentials*.

Even if cloud data are stored encrypted according to best practices, a *Key Management System (KMS) breach* can still render its security barriers ineffective. Once an attacker successfully penetrates the KMS and gained access to the keys used to encrypt and decrypt data, they can easily reconstruct all encrypted user sensor data, activity labels, identity information, etc. into plaintext form. For example, an attacker may be able to decrypt all user data by obtaining the master key used to generate and distribute data encryption keys through an internal service vulnerability or supply chain attack against the KMS. In the most extreme case, if an attacker is able to penetrate and *take full control of the cloud service provider's infrastructure*, such as by compromising its virtualization layer or the management systems of its storage clusters, they would gain the highest level of access to all tenant data. This would allow an attacker to bypass all logical isolation and security controls and directly access and steal all HAR data stored on the underlying physical

devices, regardless of their encryption status.

In addition to directly attacking core data storage, attackers may also choose to intercept data at the transmission or processing point in the data lifecycle. This is mainly manifested as **leakage at the data processing chain** in the cloud. For example, after the sensitive data reaches the cloud, they enter a complex data processing pipeline (ETL process, i.e. Extract, Transform, Load). In this pipeline, the data undergoes cleansing, feature extraction, and possibly anonymization. If an attacker is able to compromise one of the processing nodes in the data pipeline, they can intercept the stream of sensitive information containing people with disabilities' original identity and sensor data correspondences before the data have been fully anonymized or desensitized. A concrete example is when sensor data are temporarily associated with a user ID for data validation or de-duplication in a data aggregation service, an attacker may exploit a vulnerability in the service to steal the information. In addition, in order to support personalized services (e.g., personalized exercise plans or abnormal behavior alerts), the HAR system may need to temporarily de-anonymize the data at the compute instance or application layer to re-associate the sensor data with the user ID for computation. If an attacker is able to successfully compromise these servers used for data computation or service provision (e.g., by exploiting a web application vulnerability or a remote code execution vulnerability), they can obtain people with disabilities' complete sensor data, activity labels, and identity information, etc. while the data are processed in this "plaintext" state, thus enabling high-level privacy breaches.

2.1.2 Semi-Supervised Learning Attacks

Principles of Attack

In HAR systems, the threat posed by attackers reaches its highest point when they have simultaneous access to sensitive sensor data of all users and all their corresponding sensitive attribute labels, as this allows them to accurately identify, track, and infer user behaviors, posing the risk of a catastrophic privacy breach. These attackers, however, often face the challenge of obtaining large-scale, high-quality labeled data, which is costly in practice and limited by privacy regulations and protection mechanisms. As a result, attackers are rarely able to achieve the data collection required for ideal supervised learning. However, even if attackers only have access to some of the users' sensitive sensor data and the corresponding sensitive attribute labels, they can utilize this information for semi-

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

supervised learning attacks. Such attacks can be further categorized into two strategies: one is to obtain a small portion of data from all users and learn universal patterns through broad but sparse labels; the other is to obtain all the data from a few users and learn unique patterns by deeply analyzing individual subjects. Regardless of which strategy, the core of semi-supervised learning attacks lies in utilizing limited sensitive information to perform privacy inference on the entire user population, posing a serious challenge to the privacy protection of HAR systems.

In this thesis, we focus on the first attack strategy, where an attacker is able to access a small portion of all (or most) users' sensitive sensor data and their corresponding sensitive attribute labels (e.g., identity, gender, age, disability status). This attack scenario is more prevalent and stealthy in practice, as the attacker does not need to fully compromise a single user's account or cloud database, but rather builds his/her attack capability by aggregating fragmented information from a large number of users. The danger of this strategy is that it allows attackers to gradually accumulate and learn generalized association patterns among user groups without triggering widespread alarm. For example, even with access to only a few hours of activity data and disability labels for each user, an attacker may be able to utilize these sparse but wide-ranging labeled pieces of information and combine them with the large volume of unlabeled sensor data in HAR systems to train a highly accurate semi-supervised learning model, enabling the inference of users' disability attributes from other unknown data. Therefore, this attack pattern enables attackers to go beyond the limitations of existing labels. The possible attack scenarios are shown in Figure 2.3.

Attack Scenarios and Examples

Attackers could use **de-anonymization attacks** to obtain sensitive labels for all (or most) users. This type of attack infers the identity or sensitive attributes of anonymous users by correlating seemingly anonymous user sensor data or its derived behavioral characteristics in the HAR system with background knowledge or auxiliary databases in the attacker's possession that come from external data sources. The difference between the scenarios of supervised learning attacks and semi-supervised learning attacks is that, in the former, labels inferred from external databases through de-anonymization can be matched to all or most of the target records, thereby supporting supervised training. In the latter, only a subset of the target data can be linked with such inferred labels, leaving a large portion unlabeled and thus requiring semi-supervised techniques to exploit both

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

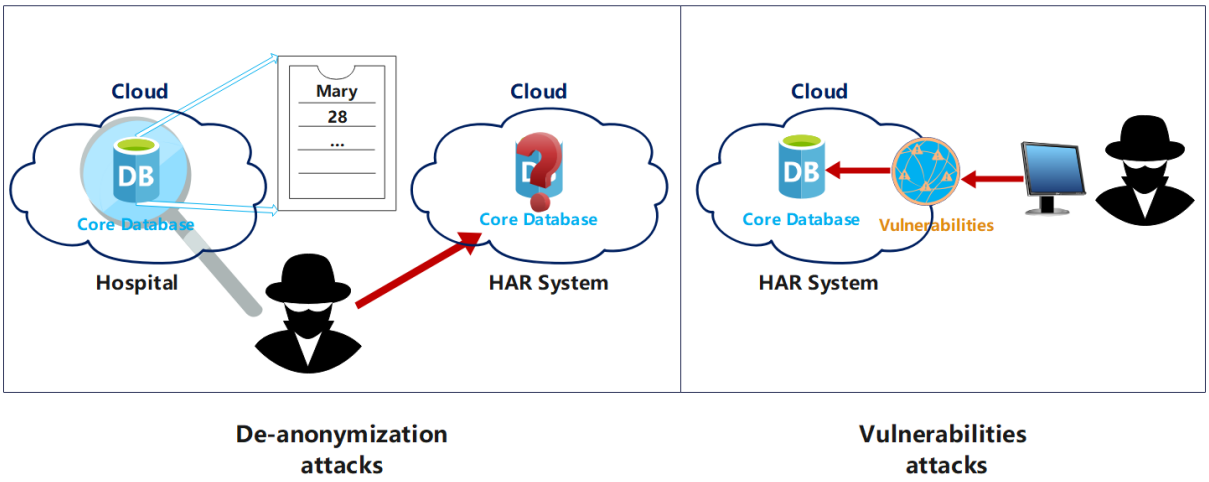


Figure 2.3 – Examples of Semi-supervised Learning attacks in the cloud.

labeled and unlabeled samples.

In the case of identity recognition, for example, an attacker may have breached a healthcare organization's database to obtain the true identity of patients. If the healthcare organization cooperates with a certain fall detection service system that collects anonymous sensor data from residents, the attacker can try to establish a correlation between the identity of the healthcare database and the anonymous user identifier of the HAR system. Since the identifiers between systems may not be perfect matches, or only short and fuzzy correspondences can be established (e.g., by comparing specific activity logs in the medical record with unique patterns in the HAR system data), attackers cannot directly obtain identity mappings for all users. However, they were able to label a small number of sensor data for most HAR users with their true identities. This limited portion of data, but labeled with true identities, is the key to training a semi-supervised learning model, which will learn the association of user behavioral patterns with identities, and thus infer user identities for the large amount of unlabeled sensor data.

For the identification of other sensitive attributes (e.g., gender, age, health status, etc.), attackers can utilize similar strategies. They can obtain anonymized sensor data from the HAR system and a small amount of volunteer data labeled with precise sensitive attributes through publicly available medical research datasets or health application databases. By comparing device characteristics or user behavior patterns in the HAR system with publicly available data, an attacker can "migrate" a small number of sensitive attribute labels to the matching anonymous user sensor data in the HAR system. If

attackers combine identity and sensitive attribute recognition, they may be able to infer other deeply sensitive information about users. For example, once a user is identified and combined with his or her inferred specific activity patterns (e.g., frequently leaving the bed at night), an attacker may further infer potential health issues (e.g., insomnia or high risk of falls), thus posing a more comprehensive privacy threat.

Another way to accomplish such an attack is to **exploit vulnerabilities** that exist in the HAR system itself. This includes *exploits of database or API vulnerabilities* as well as *cloud storage or network misconfigurations*. In this case, the attacker does not seek to obtain all sensor data, but rather exploits these vulnerabilities to query or traverse the system in a restricted manner to gain a particular sensitive attribute or attributes and its corresponding sensor data from all (or most) users. For example, in an HAR system that monitors the activities of the elderly, anonymized sensor data and corresponding health labels are stored. An attacker may be able to exploit a vulnerability in the HAR system's API by constructing a specific query that does not provide full access to the entire database but selectively obtains a small number of "disability status" labels corresponding to the anonymized sensor data, e.g., the system may incorrectly respond to a vulnerability that correlates a particular gait pattern or activity trajectory with the status of "wheelchair user". The key to this attack is that the attacker does not need access to the full database, but can obtain a portion of the sensitive attribute labels associated with the sensor data through a well-constructed query. In addition, if a cloud storage bucket storing anonymized sensitive data is publicly accessible due to a misconfiguration, an attacker can directly download these files, thereby accessing some of the user's data and labels without a sophisticated attack. As an example, if a cloud storage bucket that stores motion sensor data and anonymized disability status labels (e.g., "mobility impaired," "using crutches," "amputation") for people with disabilities is incorrectly set to public, an attacker can directly download the file containing user sensor data and their disability status labels. In these ways, attackers can obtain a small amount of labeled disability status data and then use semi-supervised learning models to infer the unlabeled sensor data to identify the disability status of other users in the system on a large scale, even if they do not have access to the complete user privacy information. However, if such sensitive attribute identification is combined with an identity recognition attack, the attacker may be aware of the user's sensitive health status, thus posing a more serious privacy and security threat.

2.1.3 Unsupervised Learning Attacks

Principles of Attack

A fundamental but equally privacy-invasive scenario in HAR systems is that an attacker can only access to a large amount of the user's sensitive sensor data (e.g., continuous data streams such as accelerations, gyroscopes, heart rates, etc.), but not to any directly corresponding privacy labels (e.g., user id, age, gender, disability status, etc.). In this scenario, known as an unsupervised learning attack, it is true that the attacker cannot directly utilize known labels to train a model to identify specific sensitive attributes as in a supervised or semi-supervised learning attack. However, this does not mean that attackers are powerless. Instead, they can use advanced unsupervised learning techniques to conduct in-depth mining and analysis of these seemingly "pure" and unlabeled data. Through methods such as clustering, dimensionality reduction, and association rule mining, attackers are able to discover hidden patterns and structures within the data. For example, they can identify user groups with similar behavioral patterns, detect abnormal behavior that deviates from the norm, and even infer implicit associations between certain behavioral characteristics and potentially sensitive attributes (e.g., gait patterns may be associated with age or disability status) without explicit labels, posing an indirect threat to the privacy of people with disabilities.

Although this threat is indirect, it poses a risk to the privacy of people with disabilities that cannot be ignored. By identifying commonalities or behavioral anomalies in a user population, an attacker can narrow down the scope of inferences, providing critical clues for subsequent targeted privacy attacks (e.g., identity inference, behavioral analysis). More importantly, unsupervised learning attacks lay a solid foundation for more sophisticated and refined privacy attacks in the future, as they enable attackers to construct user behavior profiles and classification systems without requiring prior labels, thus reducing the labeling costs and technical barrier required for subsequent supervised or semi-supervised learning attacks. The possible attack scenarios are shown in Figure 2.4.

Attack Scenarios and Examples

Because of the physical or logical separation of data and labels, or the fine-grained segmentation of cloud service role permissions, an attacker may only be able to access sensor data but not the corresponding sensitive labels. For example, a HAR system stores users' sensor data (e.g., accelerometer, gyroscope readings) in a cloud storage bucket

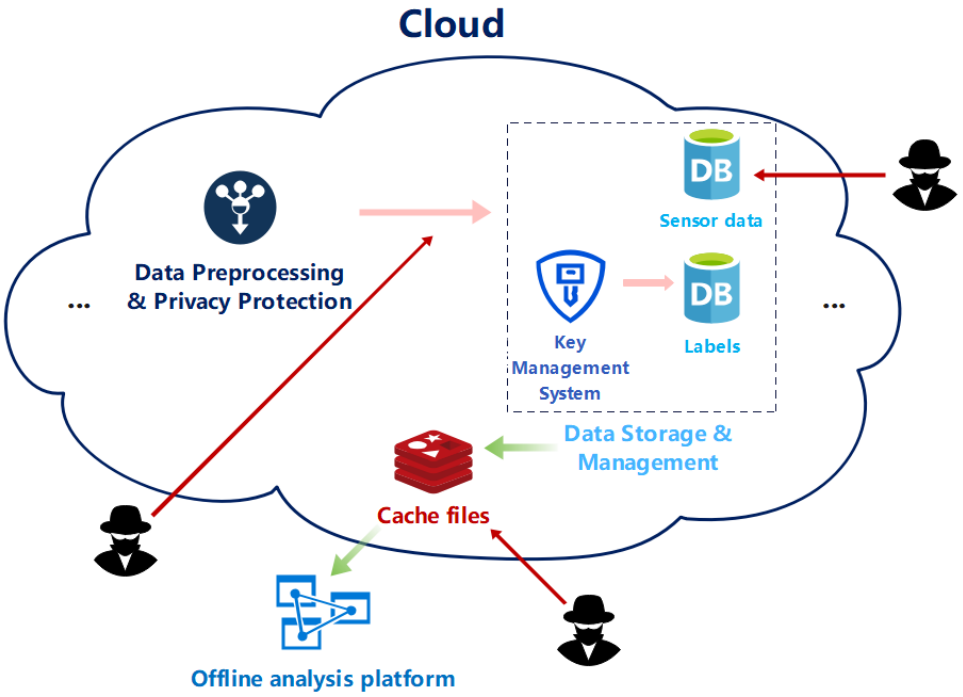


Figure 2.4 – Examples of Unsupervised Learning attacks in the cloud.

or database with a low level of privacy protection, while metadata containing sensitive labels, such as the user identity, disability status, etc. are stored in another severely restricted service. Once an attacker breaches the former (e.g., by discovering a public access vulnerability in the storage bucket), it is possible to obtain pure sensor data, such as gait, heart rate, or sleep activity sequences, for all users, but there is no direct way to know to which user the data belong or any of their corresponding sensitive attributes. Similarly, an attacker may be able to steal the credentials of a service account with specific permissions. This account may be configured to only read sensor data for all users, but not have access to user metadata services that contain sensitive labels. For example, an internal microservice account used for data backup may be granted permission to read all sensor data files, but not have access to the database that stores patient IDs and health profiles. If an attacker takes control of this account, he or she can batch download these sensor data without access to any associated sensitive labels.

Another class of attacks occurs during the lifecycle phase when labels are not associated with data. This includes accidental leakage of data after anonymization and unlabeled data export or caching. As a typical example, after anonymizing user data, HAR systems store de-identified sensor data for scientific research or analysis. If these anonymized

datasets are accidentally leaked or their storage is not configured properly, an attacker can gain access to the anonymized data of all users. Although identity information is not easily recognizable, the data may still contain behavior patterns, gait characteristics, and other information that can be used for unsupervised analysis. In addition, in some data processing workflows, sensor data are temporarily exported to a file system or cache for subsequent analysis or transmission. If these temporary export or cache files do not contain sensitive labels in order to save storage space or to simplify the process, an attacker with access to this temporary storage can only obtain these pure data. For example, the HAR service temporarily exports all user sensor data to a shared file system before transmitting the data to an offline analytics platform, without any user IDs or sensitive labels attached to the export process. An attacker who compromises this file system can then access these unlabeled sensor data.

Finally, publicly released anonymized datasets may also be the basis for unsupervised learning attacks. Certain research institutions or health platforms may publicly release strictly anonymized HAR datasets in order to promote academic research or technological development. These datasets typically remove all labels that directly and indirectly identify users or sensitive attributes, retaining only sensor data and some generic activity labels. An attacker can directly download and analyze these datasets. Although these data are designed to be anonymized, through advanced unsupervised learning techniques, an attacker may still be able to discover hidden group behavior patterns, types of abnormal activities, and even identify subgroups with similar physiological characteristics (e.g., specific gait patterns may suggest an age group) through clustering analysis. These findings, while not directly revealing individual identities, can provide a foundation for attackers to construct user behavior profiles and may offer clues for subsequent de-anonymization attacks when combined with other information.

2.2 Datasets

In this thesis, two public datasets are selected to conduct the experimental analysis: HASC-IPSC [195] and WISDM [196]. These datasets are chosen for the following reasons: 1) they both have sufficient participants as well as ample and diverse data; 2) most importantly, both of them contain subject labels, not just activity labels. HASC-IPSC also includes information about each subject's age, gender, height, and weight, which can help us to analyze the relationship between sensor data and individual characteristics;

and 3) they measured the data of various daily activities with different devices that were placed at different locations. This can better reflect the real-life scenario, because in daily life, everyone has different habits and living environments, and uses different smartphones, so they can help us better simulate real-life conditions. The experimental setup and the detailed description of these two datasets can be found in Sections 2.2.1 and 2.2.2, respectively.

2.2.1 HASC-IPSC

The HASC-IPSC dataset is designed to provide balanced and comparable data for indoor localization, floor plan generation, and activity recognition studies. The experimental data collection for this dataset was conducted at Nagoya University, specifically covering the North, Center, and South buildings. One hundred and sixteen different routes were designed for the experiment, of which fifty-six were round-trip routes within the same building, and the other fifty-eight were routes to and from different buildings. Figure 2.5 provides the target buildings for data collection and two examples of route visualization. Thus, this dataset contains six basic human activities: walking, running, ascending stairs, descending stairs, skipping, and standing still.

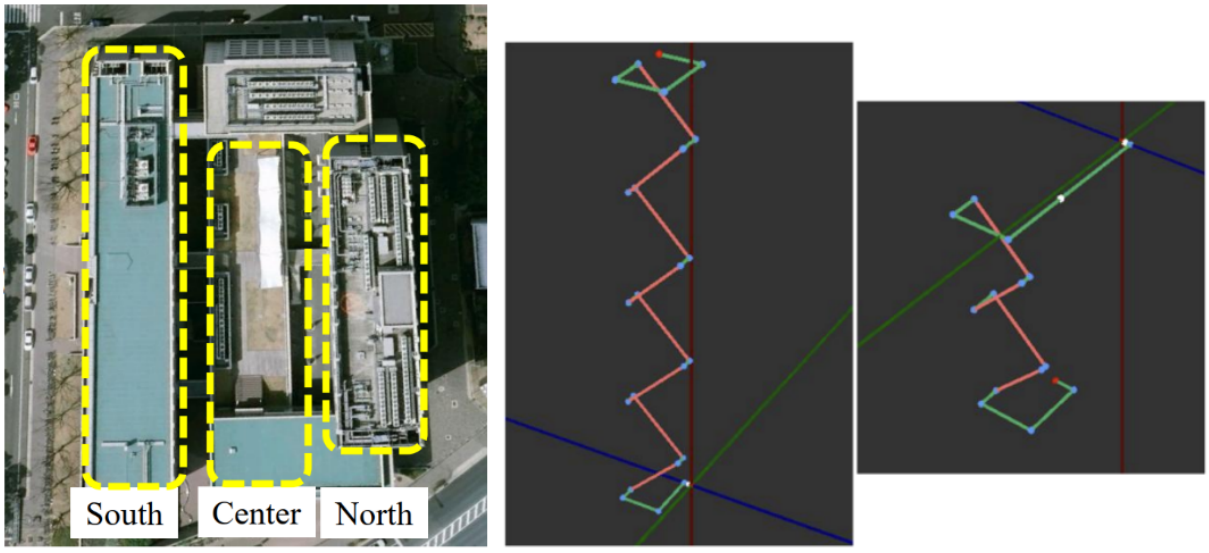


Figure 2.5 – Target buildings for data collection and Examples of route visualization [195].

A total of 107 subjects with balanced gender and age distribution participated in the data collection. The age of the participants ranged from 20 to 70 years with 4-7 males and

4-6 females in each age group. In addition, two of them were excluded due to incomplete data, resulting in a total of 55 females and 50 males. Please refer to Table 1 for more details.

Table 2.1 – Statistical information on Gender and Age of Subjects

Age	**Male**	**Female**
Early 20s	6	5
Late 20s	6	5
Early 30s	6	4
Late 30s	4	7
Early 40s	5	6
Late 40s	4	5
Early 50s	6	6
Late 50s	4	6
Early 60s	5	6
Late 60s	4	5
Total	**105**	

The experiment used Nexus 4 (Android 4.2.1) and Galaxy S3 (Android 4.1.2) smartphones for data collection. Regarding their placement, the Nexus 4 device was placed in a holder on the back of the subject's waist. The Galaxy S3, on the other hand, was carried in a normal manner, such as in a shirt pocket or bag. The HASC Logger application was installed on the devices to record sensor data. The activity data was recorded by various sensors embedded in the smartphone such as the gyroscope, accelerometer, air pressure sensor, magnetometer, light sensor and proximity sensor. In addition to various types of sensor data, the HASC-IPSC dataset also provides information on the structure of the building, the length of each link, and the precise coordinates of the route at the same time. In particular, the personal information file records detailed metadata about the subject, including person ID, gender, age, height, weight, shoe type, etc.

The data in the dataset were categorized into two main types: segmented data and sequence data. Segmented data is mainly used for training, where each data set contains a single activity with an observation time of 20 seconds. Sequence data is mainly used for testing, and contains sequences of multiple basic activities, each of which lasts at least 10 seconds, for a total observation time of 5 minutes. However, in our experiments, sequence data were used for both training and testing to better simulate real-life scenarios, as HAR systems continuously monitor user behaviors. Furthermore, since the sampling rates of

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

different sensors in the dataset vary and not every participant has data from all types of sensors, we used only the acceleration data with a sampling rate of 100 Hz in this thesis.

2.2.2 WISDM

The WISDM dataset was collected by members of the WISDM Lab in the Department of Computer and Information Sciences at Fordham University for use in HAR and biometric modeling research. The dataset recorded data from accelerometer and gyroscope sensors of a smartphone and a smartwatch, with a total of 51 subjects participating in 18 different activities of daily living. The participants consisted mainly of undergraduate and graduate students aged 18 to 25 years. For each subject, every activity was performed for 3 minutes, so each subject contributed 54 minutes of data in total.

In the experiment, each subject wore the smartwatch on their dominant hand and the smartphone was placed in their pocket. Data collection was controlled through customized apps running on the smartphone and smartwatch. The smartphone models used include Google Nexus 5/5x or Samsung Galaxy S5 (running Android 6.0 Marshmallow), and the smartwatch was LG G Watch (running Android Wear 1.5). Sensor data was collected at a sampling frequency of 20 Hz.

The dataset contains time-series sensor data from smartphone and smartwatch accelerometers and gyroscopes for a total of four sensors (phone accelerometer, phone gyroscope, watch accelerometer, watch gyroscope). All time-series data were labeled not only with the activity performed, but also with subject identifiers, which allows the data to be used for building and evaluating biometric models, as well as activity recognition models. The dataset covers 18 diverse activities of daily living, as shown in Figure 2.6, including basic walking activities (e.g., walking, jogging, climbing stairs), manual routines (e.g., brushing teeth, folding clothes) and various eating activities (e.g., eating pasta, eating chips). These activities were further categorized into three groups: *non-hand-oriented* activities, *general hand-oriented* activities, and *hand-oriented* activities *related to eating*, as shown below. Therefore, this dataset contains various life scenarios. In our study, given the limitations of our experimental environment, only the acceleration data of the smartwatch were used. This also ensures consistency with the other dataset.

— **Non-hand-oriented activities:** walking, jogging, stairs, standing, kicking

— **Hand-oriented activities (General):** dribbling, playing catch, typing, writing, clapping, brushing teeth, folding clothes

— **Hand-oriented activities (eating):** eating pasta, eating soup, eating sandwich, eating chips, drinking

Activity	Code
Walking	A
Jogging	B
Stairs	C
Sitting	D
Standing	E
Typing	F
Brushing Teeth	G
Eating Soup	H
Eating Chips	I
Eating Pasta	J
Drinking from Cup	K
Eating Sandwich	L
Kicking (Soccer Ball)	M
Playing Catch w/Tennis Ball	O
Dribblinlg (Basketball)	P
Writing	Q
Clapping	R
Folding Clothes	S

Figure 2.6 – The 18 Activities Represented in Data Set [196].

2.3 Conclusion

In summary, this chapter provides a detailed definition of the attacker model employed in this thesis and the datasets used. The adversary is assumed to have substantial computational resources and expertise in ML/DL, and aims to perform identity inference attacks on time-series data collected from individuals with disabilities via cloud services. By training ML or DL models, the attacker can learn patterns and correlations between the data and user identities, enabling the inference of sensitive information. Attackers may vary in their capabilities, including the amount of data they can access and the proportion of labeled information available, which directly affects the accuracy and success of their attacks. The consideration of these variations enables a more realistic assessment of potential threats and the effectiveness of privacy-preserving measures.

In addition, an exhaustive description and analysis of the datasets utilized for experimental validation are provided. These datasets can help us simulate real-world attack scenarios. However, currently there are no publicly available HAR datasets that include persons with disabilities, even though their privacy is at the core of our motivation. Therefore, we relied on two existing datasets that offer relatively large numbers of participants, diverse activities, and a reasonable balance across participants (e.g., in terms of gender). Although the experimental data does not contain disability information (i.e., the datasets do not include participants with disabilities), the sensitive characteristics they exhibit are sufficient to reflect the fundamental privacy issues faced by such data, constituting the core challenge that our research aims to address. Furthermore, in this thesis, we also constructed a new dataset that explicitly includes participants with disabilities and incorporates activities that are closer to real-life scenarios (e.g., daily household tasks), while exploring whether time-series data can further distinguish between able-bodied and disabled individuals. This dataset will be presented in Chapter 5 as an initial attempt that lays the foundation for subsequent research.

After clarifying the threats posed by advanced inference attacks and analyzing the characteristics of the target data, we argue that traditional de-identification techniques are no longer sufficient to provide robust protection. These conventional methods often fail to prevent attackers from exploiting subtle correlations in the data to infer sensitive information. Consequently, the following sections build on the attacker models established in this chapter to develop and propose an innovative privacy-preserving solution. The proposed approach aims not only to mitigate the risks of identity inference but also to maintain data utility, ensuring that the data remain accessible and useful for legitimate analytical purposes. By balancing strong privacy protection with data availability, this work seeks to protect the sensitive information of individuals with disabilities against increasingly sophisticated adversaries.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

Chapter 3

Privacy-Preserving Feature Selection in Local Environment

In assisted living scenarios, the use of sensors embedded in wearable devices and smartphones for HAR has become a key technology in smart monitoring systems and shows great potential for application. Compared to audio or video data, time-series data from sensors such as Inertial Measurement Units, accelerometers, and gyroscopes, offers lower computational energy consumption, and greater environmental adaptability, making it more universally applicable in daily monitoring. Specifically, HAR technology based on time-series sensor data can accurately identify and record user daily behavior patterns, such as wheelchair use, standing up, and limb movements, providing quantitative evidence for evaluating rehabilitation training outcomes and improving independent living abilities. Additionally, in the event of an emergency (such as a fall), the system can respond immediately and automatically, and send alerts to caregivers or medical institutions.

The relationship between these application scenarios and the data flow architecture of the intelligent monitoring system is shown in Figure 3.1. During the local phase, the wearable devices or smartphones collect biological signals such as heart rate, as well as inertial sensor data, to generate rich time-series data streams. These data are preliminarily processed locally and then securely transmitted to the cloud via transmission layers such as Bluetooth, Wi-Fi, or cellular networks. In the cloud, powerful computational resources enable AI algorithms and pattern recognition on large-scale data, accurately identifying specific activities or abnormal events. The processed results are available for in-depth analysis by professional institutions such as hospitals and rehabilitation centers, providing personalized health management and emergency assistance services for people with disabilities, effectively enhancing their quality of life and safety.

In sensor-based HAR research, processing raw, high-dimensional time series data is a core challenge. To overcome this problem, researchers typically employ feature extraction and feature selection techniques aimed at reducing the dimensionality of large-scale data

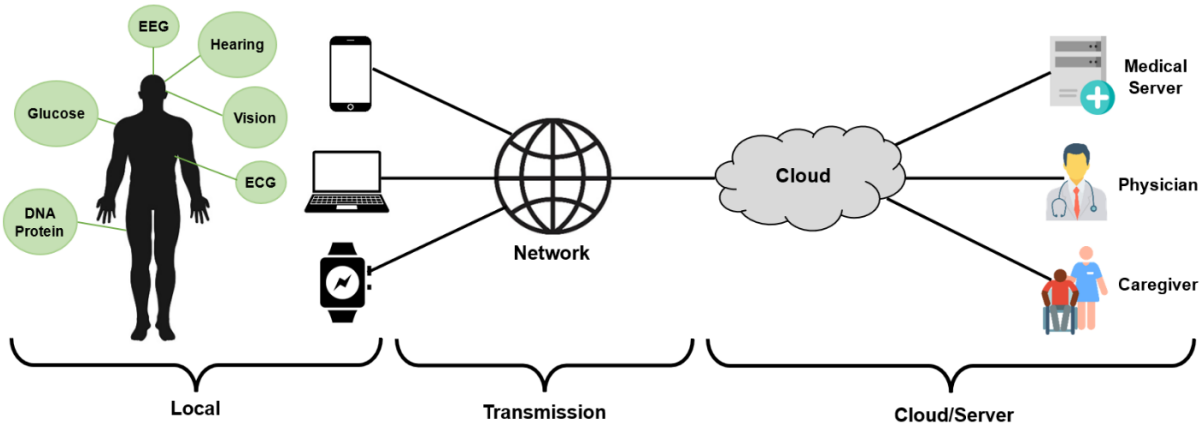


Figure 3.1 – The relationship between time-series data application scenarios and the data flow architecture of the smart monitoring system.

and obtaining the most representative feature sets, such as [197, 198, 199]. This process not only effectively improves the accuracy of classifiers, but also shortens model training time and mitigates overfitting issues, thereby constructing efficient HAR models. However, this dimension reduction optimization framework often overlooks a critical privacy risk: while the extracted and selected features may ensure or even enhance activity recognition performance, they may inadvertently leak users' sensitive personal information. For example, attackers could exploit these seemingly harmless features to infer user personal attributes such as gender, age, height, weight [107], or even identify specific users to whom the data belongs [200, 201, 202], thereby posing potential threats to user identity, etc. Therefore, this chapter not only focuses on model recognition accuracy, but also delves into how to effectively mitigate privacy risks caused by feature data leakage while maintaining activity recognition performance, achieving a balance between performance and privacy.

Given the challenges faced by the HAR model in terms of performance optimization and user privacy protection, researchers have begun to explore a variety of solutions. These solutions aim to address this fundamental contradiction at different levels, ranging from data processing to model inference. According to the taxonomy introduced in Chapter 1 (local, transmission, and cloud), it is worth noting that although overall solutions may span multiple levels, the privacy-preserving modules are typically implemented at the local level, ensuring that sensitive data is protected before transmission. As dis-

cussed in Chapter 1, the solutions for protecting the privacy of time series data include collaborative operations, data sanitization and federated learning. Table 3.1 presents a comparison between state-of-the-art privacy-preserving approaches for time-series data and our proposed method. In evaluating the Computation aspect, we specifically consider the feasibility of each method on IoT devices relevant to HAR, reflecting the typical resource constraints of wearable sensors and mobile devices. For the Privacy level, our assessment incorporates three key factors: the inherent privacy protection mechanism of the method, the likelihood of data leakage during transmission or storage, and the potential impact on sensitive information if such a leakage occurs. This approach ensures that privacy evaluations account not only for theoretical guarantees but also for practical risks in real-world HAR scenarios.

Table 3.1 – A Comparative Study of Privacy Protection Methods for Time Series Data.

Method	**Introduced bias**	**Computation**	**Level of privacy**
HE + LDP	Moderate (noise from LDP reduces accuracy)	Very High (HE heavy; often infeasible for IoT)	Very High (LDP + encryption; strong, but key risk remains)
RG-RP + LSTM-CNN	Small (minor accuracy drop)	Low (only RG-RP on device; LSTM-CNN runs on server/edge)	Medium (random grouping / projection, no formal DP)
Data Sanitization	Small (slight utility loss, sensitive inference blocked)	Low (training heavy off-device; inference light on-device)	High (sensitive signals replaced via autoencoder)
FL + HE	Minimal (encryption adds no statistical bias)	Very High (local training + HE heavy for IoT)	High (encrypted updates; strong crypto but key risk remains)
FL + LDP	Moderate (noise impacts accuracy depending on ϵ)	High (local training heavy; noise addition light)	High (model updates perturbed with formal LDP)
Privacy-Preserving Feature Selection (Our proposed method)	Small (patterns preserved, limited accuracy loss)	Minimal (local feature extraction light; selection on edge server)	High (sensitive features filtered before sharing)

Among the methods considered (as shown in Table 3.1), the HE- and FL-based approaches are not suitable for local IoT environments (i.e. local devices) due to their high computational overhead. In addition, the RG-RP approach provides a relatively limited level of privacy protection. Therefore, data sanitization can be regarded as a more ap-

propriate state-of-the-art solution in this context. However, data sanitization has notable limitations: it typically requires a relatively large dataset to train the autoencoder effectively, which may demand higher computational resources and stronger servers; performing training on edge servers is possible but may reduce model generalization. In addition, the processed data transmitted after sanitization can still be relatively large, imposing additional load on the IoT devices (i.e. the communication cost). In contrast, our approach offers a lightweight and efficient alternative, as only the non-sensitive feature subsets are extracted locally (i.e. on the devices) with minimal computation before being transmitted. At the same time, reduced information content lowers the potential risk of privacy leakage. This makes it particularly well-suited for resource-constrained HAR devices while maintaining privacy protection.

In our approach, the trade-off between privacy and accuracy is defined as a multi-objective optimization problem. Our method aims to identify the smallest possible subset of features that ensures high accuracy in activity recognition while maximizing user privacy protection. Compared to traditional privacy protection strategies (e.g., applying perturbation schemes to raw data), the core advantage of our method lies in its noise-free nature. The smart monitoring system transmits only those features that are most representative of activity recognition, which are carefully selected to ensure classification performance while preventing reverse inference of sensitive personal information. This avoids potential performance degradation caused by adding perturbations, thereby achieving privacy protection while ensuring the efficiency and reliability of the HAR model.

In addition, we note that the most representative feature sets may not be static but vary depending on the user group. Considering significant differences in behavioral patterns and physical conditions among different users (such as the elderly and people with disabilities), how to dynamically provide and update the corresponding feature subsets for different user groups to ensure adequate privacy protection and maintain identification accuracy is also one of our concerns. To this end, we propose a deployment architecture based on cloud-edge collaboration. The core idea of this architecture is that a centralized and more powerful edge server (such as one deployed at the neighborhood level) is responsible for dynamically selecting and updating the most discriminative feature subsets based on the collected large-scale data. Local devices, such as wearable devices or smartphones, are only responsible for performing a lightweight task: extracting the corresponding features based on the feature subset specified by the edge server and transmitting them to the HAR system server. This model not only significantly reduces the risk of privacy leakage

and communication costs by minimizing the amount of data transmitted (only extracted features are transmitted), but also enables the system to self-adjust according to changes in user groups, achieving an efficient, personalized, and privacy-aware HAR model.

In this chapter, we mainly consider the supervised learning attack model described in Chapter 2. The attack model assumes that the attackers have access to all sensitive user data and the corresponding sensitive attribute labels. Our defense goal is to ensure that even if anonymous user identity information is unfortunately leaked, attackers cannot accurately infer the user identity of the owners of newly added data using existing data or learned models. To achieve this, the Multi-Objective Evolutionary Algorithm (MOEA) is utilized to explore the feature subsets that can accurately recognize human activities while limiting user identification.

3.1 Multi-Objective Feature Selection

In this section, our objectives will be explained first. Then, we will introduce our privacy-preserving feature selection approach and all the classification models used in the proposed method. Figure 3.2 shows the whole process of selecting the best feature subset. First, the framework collects multi-dimensional time series data from users via wearable devices. Then, a comprehensive feature set is extracted from this data, including features from multiple dimensions (i.e., the statistic, temporal, and frequency domains). Next, a multi-objective feature selection algorithm is used to select and optimize an initial subset from the large feature set based on the predefined objectives. Finally, several classification models are used to validate the final optimal feature subset, i.e., testing activity recognition and user recognition accuracy.

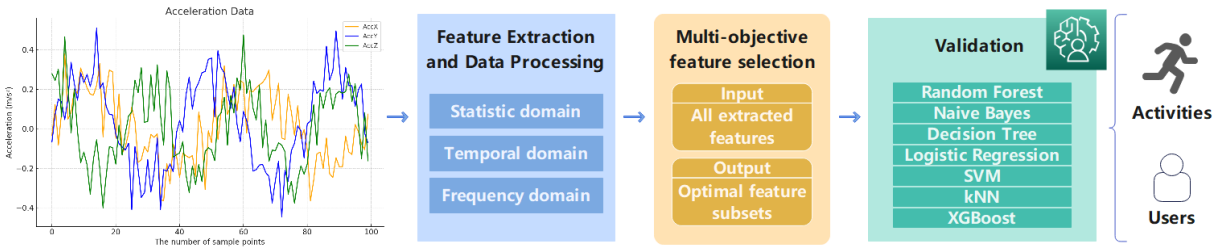


Figure 3.2 – System architecture of the multi-objective feature selection.

3.1.1 Objectives of Privacy-Preserving Feature Selection

Our work aims to identify a feature subset that can achieve high activity recognition accuracy while concealing user identity, and that is suitable for local feature extraction on low-power devices. The feature selection process can be implemented on edge servers, such as those in healthcare institutions. To this end, in addition to objectives related to increasing HAR accuracy and reducing user recognition accuracy, the objective of reducing the number of features is proposed to improve algorithm convergence, thereby significantly reducing the number of generations required, making our method suitable for resource-constrained systems (i.e., the edge server).

For better illustration, the set of all features is defined as F and the selected subset of features is defined as f. In the Multi-Objective Feature Selection algorithm, using the same input feature subset f, two classifiers are trained: the first M_1 is for recognizing human activities, and the second M_2 is for identifying the users. The recognition accuracy of each classification model is defined as: $Acc = \frac{1}{K}\sum_{i=1}^{K} \mathbb{1}(y_i, \hat{y}_i)$, where K is the total number of samples, y_i is the true label of the i-th sample, $\hat{y}_i$ is the predicted label of the i-th sample, and $\mathbb{1}(y_i, \hat{y}_i)$ is an indicator function that equals 1 if $y_i = \hat{y}_i$, and 0 otherwise. Therefore, our objectives used in MOEA are defined as follows.

— The first objective O_1 is to maximize the accuracy of the HAR task on a subset $f \subset F$:

$$O_1 = max(Acc_h) \tag{3.1}$$

— The second objective O_2 is to minimize the accuracy of user recognition model on the same subset $f \subset F$:

$$O_2 = min(Acc_u) \tag{3.2}$$

— To further enhance privacy protection, the third objective O_3 is defined as maximizing the performance difference between the HAR and user identification models:

$$O_3 = max(Acc_h - Acc_u) \tag{3.3}$$

— The fourth objective O_4 is to minimize the number of selected features f. This not only accelerates the convergence of the feature selection process on the edge server, but also reduces the computational burden on IoT devices (since fewer features need to be calculated once the selection is made) and decreases the bandwidth required

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

for subsequent data transmission.:

$$O_4 = min(\frac{|f|}{|F|}) \tag{3.4}$$

3.1.2 Multi-objective Optimization for Feature Selection

Multi-objective evolutionary algorithms (MOEAs) are a class of optimization methods designed to simultaneously optimize multiple, often conflicting, objective functions. In many problems, improving one objective may lead to worse results for others [203], making it impossible to find a single "best" solution. Instead, MOEAs aim to provide a set of high-quality solutions that represent different trade-offs among the objectives.

The core principle of MOEAs is inspired by natural evolution. The process [204] typically begins with the random generation of a population of candidate solutions. Each solution in this population is evaluated using one or more fitness functions, which measure how well it performs according to the different objectives. Based on these fitness scores, solutions are selected for reproduction: better-performing solutions have a higher chance of being chosen to generate the next generation. To introduce diversity and explore the solution space, evolutionary operators such as crossover (combining parts of two solutions) and mutation (randomly altering a solution) are applied. This process of selection, variation, and evaluation is repeated over many generations, gradually improving the population towards better trade-offs among the objectives.

The resulting set of solutions is analyzed in terms of the Pareto-optimal front, which consists of solutions where no objective can be improved without compromising another [205]. The Pareto Front illustrates the trade-offs between conflicting objectives and helps identify the best possible compromises. By providing a diverse set of solutions rather than a single point, MOEAs allow decision-makers to explore multiple options and select solutions that best balance competing criteria, making them particularly effective for complex, real-world optimization problems.

In our study, the Non-dominated Sorting Genetic Algorithm III (NSGA-III) [205] is applied to explore the optimal feature subsets. NSGA-III (see Figure 3.3) is an extension of NSGA-II [206], primarily designed to enhance the diversity of solutions and adaptability to high-dimensional objective spaces. While NSGA-II works well for problems with only a few objectives (typically 2-3), NSGA-III introduces a reference point-based selection mechanism that enables better exploration of the solution space to maintain a well-distributed Pareto-optimal front in multi-objective optimization problems with 3 to

15 objectives. NSGA-III continues the concept of non-dominated sorting, dividing the population into multiple non-dominated fronts. It replaces crowding distance with reference point allocation to select the next generation of individuals, ensuring the diversity and representativeness of solutions. At the same time, NSGA-III employs genetic operations (i.e., crossover and mutation) to generate new sets of solutions, resulting in a more uniform and comprehensive Pareto front [205]. In our experiments, the Pymoo library [207] is utilized to implement the NSGA-III algorithm.

According to Figure 3.3, the process of the NSGA-III algorithm can be described as follows. The NSGA-III algorithm begins with an initialization phase, where not only is the initial population created, but also a set of uniformly distributed reference points is generated. These reference points serve as anchor points to guide the distribution of population diversity. Subsequently, the NSGA-III algorithm constructs the first generation of the population through genetic operations such as selection, crossover, and mutation, as well as non-dominated sorting. In subsequent evolutionary iterations, the NSGA-III algorithm merges the current parent generation with offspring individuals generated through genetic operations. Next, the merged population is divided into different non-dominated levels through fast non-dominated sorting. To form a new parent population, the NSGA-III algorithm sequentially selects non-dominated fronts until the inclusion of the last front causes the population size to exceed the predefined scale. At this point, the NSGA-III algorithm does not rely on the crowding distance like the NSGA-II algorithm, but instead adopts a reference point-based selection strategy: individuals are associated with their closest reference point. Subsequently, to precisely fill the remaining slots in the new parent population, preference is given to individuals associated with the reference points that have the fewest associated individuals, thereby ensuring a uniform distribution of the population in the objective space. The newly selected parent population undergoes genetic operations again to generate offspring, entering the next generational cycle. This process repeats until the maximum number of generations is reached, at which point the NSGA-III algorithm terminates.

To achieve our goal of selecting privacy-preserving features, a fitness function needs to be defined to evaluate the quality of solutions and to provide an optimization direction for the population by calculating objective values. It is important to note here that the population for the fitness function is a subset of features f. The fitness function (see Algorithm 1) provides the objective values of the solutions, which the NSGA-III algorithm utilizes for non-dominated sorting [205].

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

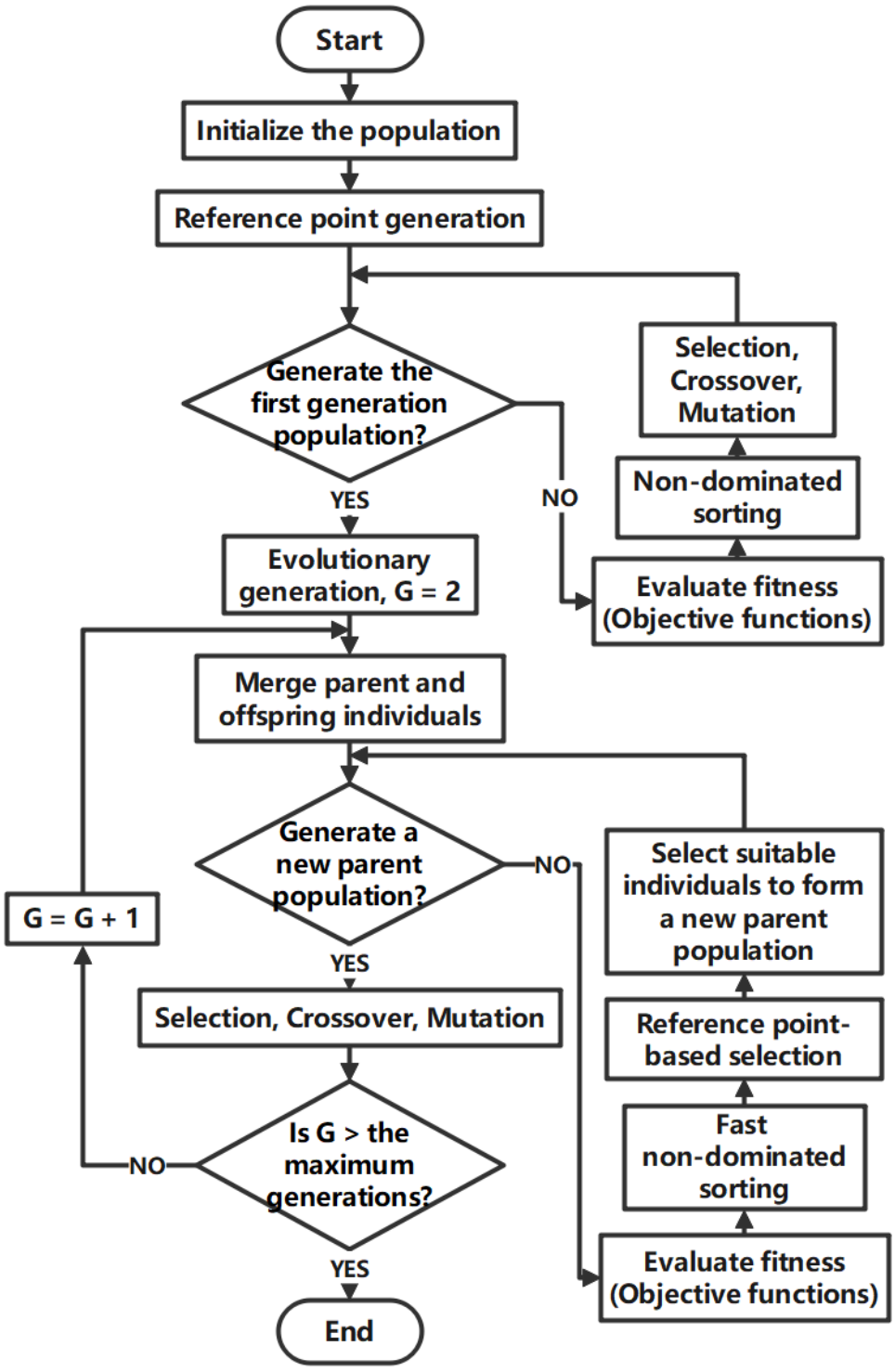


Figure 3.3 – The flowchart of NSGA-III Algorithm.

Algorithm 1 Fitness Function

procedure FITNESS FUNCTION
    Train the HAR model M_1
    Train the User identification model M_2
    Calculate the accuracy of M_1: O_1 (see Eq. 3.1)
    Calculate the accuracy of M_2: O_2 (see Eq. 3.2)
    Calculate the performance disparity between M_1 and M_2: O_3 (see Eq. 3.3)
    Calculate the normalized number of selected features: O_4 (see Eq. 3.4)
    return $-O_1, O_2, -O_3, O_4$ (*minimize* function is used in NSGA-III)
end procedure

3.1.3 Classification Models

Random Forest (RF) is used as a classifier in NSGA-III because its use is well documented in the literature [203, 208, 202]. RF is a powerful ensemble learning algorithm based on Decision Tree (DT), offering superior classification performance and generalization ability compared to DT. It reduces prediction error by leveraging the diversity among multiple trees, i.e., majority voting, and can overcome overfitting issues [209]. The advantages of RF include ease of interpretation, reasonable computing cost, and its ability to provide feature importance metrics, such as Gini importance or permutation importance, which simplify the identification of key features, making it especially useful for high-dimensional datasets [202].

In addition to RF, we also evaluate the effectiveness of the features selected by the NSGA-III feature selection system using Naive Bayes (NB), DT, Logistic Regression (LR), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN) and XGBoost (XGB). In other words, we utilize these ML algorithms to identify activities and users, respectively, in order to verify whether the selected features can accurately recognize human activities while protecting user identity. Detailed parameters for these classifiers are provided in the next section.

3.2 Performance Evaluation

This section will first introduce the feature extraction process and the initial features related to the changes in the acceleration signal, then discuss the experimental setup, and finally present the experimental results and analyze them to demonstrate the effectiveness of our feature selection method. In addition, to validate our proposed approach,

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

this chapter solely utilizes the HASC-IPSC dataset, which was previously introduced in Chapter 2.

3.2.1 Feature Extraction

Since our work focuses on selecting features that can both protect user identity and ensure HAR accuracy, feature extraction is an important first step. In this work, we use the accelerometer signals that are sampled in a fixed-width sliding window of 2.56 seconds (i.e., 256 readings, since the sampling frequency of the HASC-IPSC dataset was 100 Hz) with 50% overlap between two adjacent windows. The overlap is intentionally designed to reduce information loss during data segmentation. In this chapter, accelerometer data are used for both activity and individual recognition, as accelerometer signals effectively capture patterns of different activities and individual differences, while avoiding the additional complexity of combining them with gyroscope data. Based on [198, 201, 210], features are selected from statistical domain, temporal domain and frequency domain. The features are extracted from each window by calculating variables from the three different domains listed in Table 3.2.

Table 3.2 – Extracted features for feature selection.

Domain	**Features**
Statistical	Min, Max, Mean, Median, Standard deviation, Variance, Variation coefficient, Kurtosis, Skewness, Root mean square, Interquartile range, Median absolute deviation, Peak amplitude, Signal magnitude area, Percentiles p50
Temporal	Area under the curve, Mean absolute difference, Mean difference, Median absolute difference, Medain difference, Sum of absolute difference, Absolute Energy, Entropy, Peak to peak distance, Number of maximum peaks, Number of minimum peaks, Slope, Zero crossing rate, Autocorrelation, Axes correlation XY, Axes correlation XZ, Axes correlation YZ, Angle XY, Angle XZ, Angle YZ
Frequency	Fundamental frequency, Maximum frequency, Median frequency, Mean frequency, Spectral centroid, Spectral roll off, Spectral slope, Spectral kurtosis, Spectral skewness, Spectral entropy, Wavelet entropy

Statistical-domain feature extraction focuses on analyzing raw sensor signals to derive meaningful statistical information. Typical features include values like maximum, minimum, mean, standard deviation, variance, and root mean square. Additionally, more complex features such as skewness and kurtosis can be used to gain deeper insights into

the probability density function of the signal, allowing for analysis of its shape and distribution [211]. Temporal-domain features capture the longer-term dynamics of a signal and reveal the relationships between signals within a sequence. On the other hand, frequency-domain features are useful for assessing the distribution of signal energy and are particularly effective in identifying repetitive patterns or activities [212]. Time-domain (including statistical-domain and temporal-domain) features are commonly used due to their low computational cost. Although frequency-domain features are more computationally intensive, they provide valuable context-aware insights into the signal patterns associated with different human activities [213].

Among the features listed in Table 3.2, except for the correlation features (i.e., Axes correlation XY, Axes correlation XZ, Axes correlation YZ, Angle XY, Angle XZ, Angle YZ), which are used to measure the correlation between signals from different axes, the other 41 features are extracted separately from the x, y and z axes of the acceleration data. In other words, the total number of features extracted is $41 \times 3 + 6 = 129$. These features are extracted mainly with the help of the existing Python libraries Tsfel [214] and NumPy [215]. Our feature extraction process ultimately results in a multidimensional and discriminative feature set, which is crucial for exploring feature subsets that ensure high HAR accuracy while simultaneously reducing the user recognition accuracy, within our privacy-preserving feature selection framework.

3.2.2 Experimental Setup

All experiments are performed on a Linux machine with 100GB of RAM, a 32 core Intel Xeon E5-2640 v3 CPU and a NVIDIA RTX A5000 GPU. All models are run in Python 3.10 environment. The versions of the main packages used are as follows: *Pymoo* 0.6.1.3, *Tsfel* 0.1.6, *Numpy* 1.26.4, *Scikit-learn* 1.5.2, *Tensorflow* 2.18.0.

In our experiments, 129 features are extracted from each time window of the HASC-IPSC dataset to constitute a new initial feature set, see the previous section for details. Subsequently, these data are randomly divided into two groups: 70% of the data are selected for training (of which 20% are used as the validation set) and 30% of the data are used as testing. The optimal features are chosen according to their performance on the validation set. The final evaluation is performed on the test set. We then normalize the data in each column by transforming it to have a mean of 0 and a variance of 1. This step ensures that the features have a common scale before utilizing them in classification tasks.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

The key parameters of the NSGA-III algorithm are as follows: The population size is 455 and the number of generations is 120. In this study, NSGA-III was applied to a four-objective optimization problem ($M = 4$). The reference point division parameter $p = 12$ defines the granularity of the reference point grid on the unit simplex of the objective space, i.e., each axis is divided into 12 segments. Based on this parameter, the total number of reference points is calculated as $H = \binom{p+M-1}{M-1} = 455$. To fully utilize all reference points and maintain solution diversity, we selected a population size of $N = 455$, which is the minimum feasible value for running the algorithm. Additionally, preliminary experiments indicated that fewer than 100 generations resulted in insufficient convergence of the Pareto front, while increasing the number of generations beyond 120 showed no significant improvement. Therefore, 120 generations were chosen to balance convergence and solution diversity. The initial population is generated using BinaryRandomSampling. The Simulated Binary Crossover [207] with default parameters ensures effective exploration by combining parent solutions. Mutation is handled by Polynomial Mutation [207] with a probability of 0.1 and an eta of 20, introducing controlled variation to prevent premature convergence. Finally, duplicate solutions are removed to maintain population diversity throughout the evolutionary process.

For the classification models, the default parameters are used for NB and DT. RBF kernel is used in SVM with C = 10 and gamma = *'scale'*. The parameters chosen for the other classifiers are referenced in [216].

3.3 Experimental Results

3.3.1 Comparing different definitions of objectives

Our goal is to propose a privacy-preserving approach that can be deployed in the local environment (i.e., prior to data transmission) to address the various privacy threats that data may face during transmission and within application. In our case, feature extraction is performed on local devices, while feature selection is executed on the edge server, thereby combining local computation with edge-level processing. In [202], Nguyen et al. proposed to achieve privacy-preserving feature selection with three objectives (i.e. O_1, O_2, O_3), but this poses difficulties in deployments within resource-constrained environments. In this experiment, we will compare their fitness function with ours on the HASC-IPSC dataset. More precisely, we will confirm the necessity of defining a fourth objective O_4,

i.e., reducing the number of selected features, in achieving the task of maximizing HAR while protecting the user's identity (in resource-constrained environment such as the edge server). Figures 3.4 and 3.5 respectively illustrate the accuracy trends of HAR and user identification across generations using features selected by the NSGA-III algorithm. The distinction lies in the number of objectives used: Figure 3.4 presents results with three-objective feature selection (i.e., O_1, O_2 and O_3) as in [202], whereas Figure 3.5 showcases results with four-objective feature selection (i.e., O_1, O_2, O_3 and O_4). All other parameters in the experiment are kept consistent. Moreover, it is worth mentioning that in each generation, the optimal solution is selected according to the most important objective O_3, i.e., maximizing the difference between the two classification tasks. For clarity, only distinct results are retained in sequential order across generations, so the final generation shown in the figures is not 120.

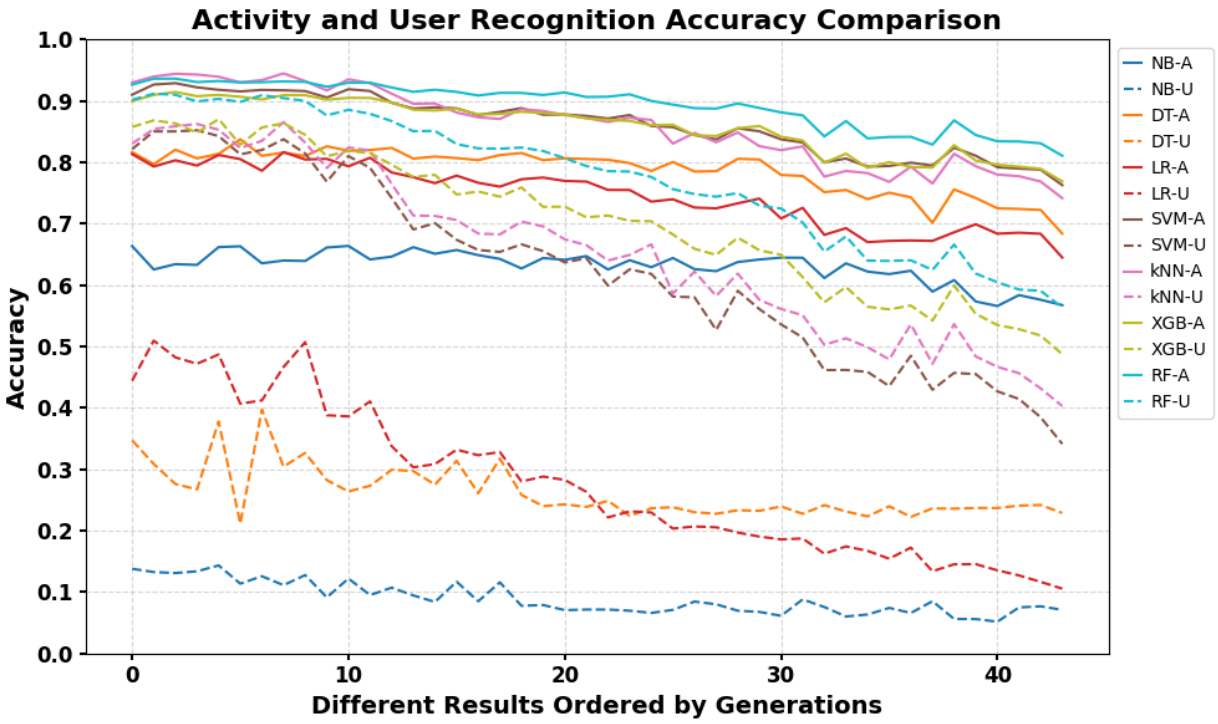


Figure 3.4 – HAR and User identification results using features selected by NSGA-III with three objectives (A indicates activity recognition accuracy, U indicates user recognition accuracy).

Since the classifier used in the NSGA-III algorithm is RF, all subsequent comparisons are based on the RF classification results (cyan line in the figure). As can be seen in

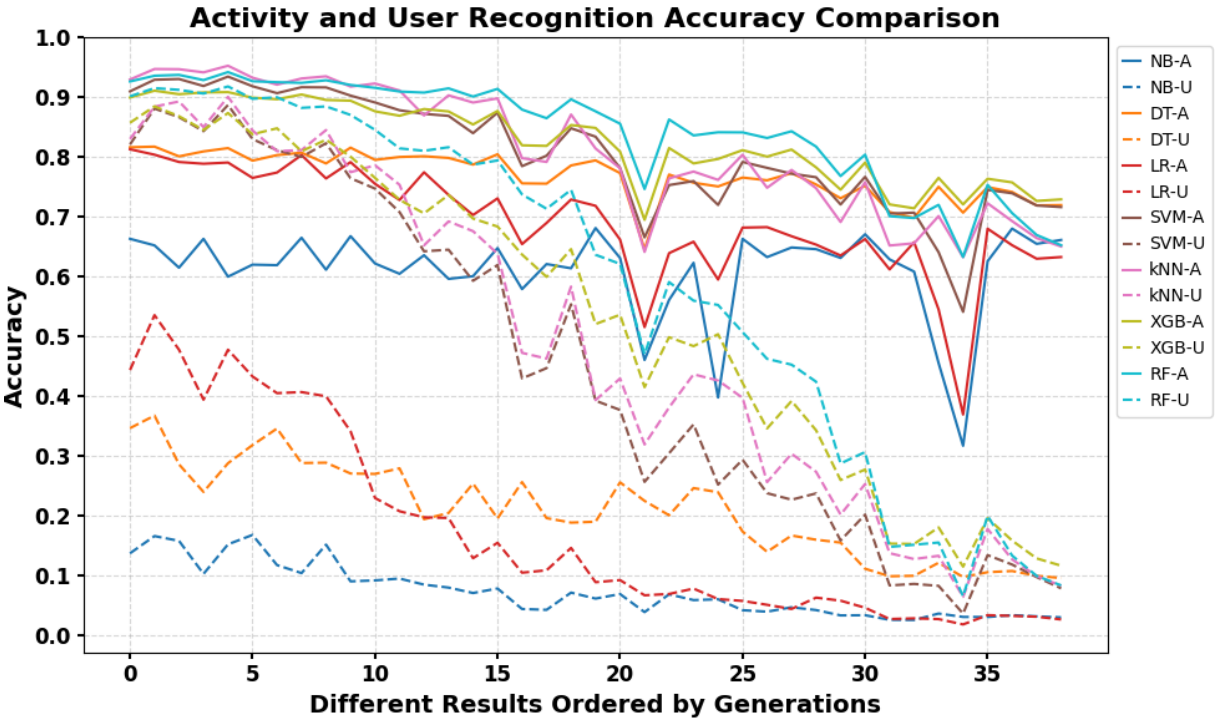


Figure 3.5 – HAR and User identification results using features selected by NSGA-III with four objectives (A indicates activity recognition accuracy, U indicates user recognition accuracy).

Figure 3.4, under the same conditions, the accuracy of user recognition is still 56.37%, even for the best result using the Nguyen et al. approach [202]. Although it has dropped significantly from the initial 94.88%, our goal is to rapidly reduce it to a random level (i.e., 50% or even lower). To achieve this, we introduced a fourth objective O_4 into the fitness function. The results are shown in Figure 3.5. Starting from approximately the 25th solution, the user recognition accuracy drops to a randomized level, eventually reaching about 10%. Therefore, compared to [202], our method can find a suitable solution more quickly. However, we will not select feature subsets beyond the 30th solution. Even though these feature subsets effectively conceal user information, they also result in a significant loss in HAR. Thus, the more desirable choices are the feature subsets obtained from solutions 25 to 30. Here, the 27th-solution feature subset is selected as the best result of this approach, and the selected features are: *fundamental frequency*, *mean frequency* of the x-axis, *std*, *variation coefficient*, *peak to peak distance*, *fundamental frequency* of the y-axis, and *fundamental frequency* of the z-axis. This selected feature subset corresponds

to the 52th generation of the NSGA-III algorithm, and it took approximately 5.9 hours to obtain this result in our test environment. In contrast, the approach proposed in [202] did not find any solution satisfying our preset conditions within the specified evolutionary generations (i.e., until the algorithm terminated). Its best solution can only be regarded as that of the final 120th generation (with 18 selected features), which required about 15.2 hours to obtain. Furthermore, we extended the number of generations for the three-objective scenarios to determine how many are required to obtain solutions meeting our criteria. The NSGA-III results stabilized around generation 130 and showed no further improvement up to 800 generations, which was the maximum tested. While adjusting parameters might alter the outcome, under the current settings identical to those used for the four-objective case, no solutions satisfy the target criteria. These results demonstrate that our proposed method achieves a significantly faster convergence speed compared to the method in [202], making it more suitable for deployment in resource-constrained edge computing environments. This advantage is of great practical significance in real-world applications.

To better demonstrate the effectiveness of our privacy-preserving feature selection method and also for comparison purposes, the results of different classifiers using all 129 features in the HAR and user identification tasks, as well as the recognition results of the optimal feature subset for both schemes, are shown in Table 3.3. As can be seen, the four-objective method has significantly better results. We define the differences across all classifiers as

$$Mean_d = \frac{1}{N}\sum_{i=1}^{N}[Acc_h(M_1^i) - Acc_u(M_2^i)] \tag{3.5}$$

where N is the number of classifier types.

If the means of the differences across all classifiers on the two recognition tasks are calculated (see Eq. 3.5), the result is 39.71 for the three-objective scheme, while it is 52.29 for the four-objective scheme. Therefore, the four-objective scheme achieves a 12.58 points improvement in privacy protection performance compared to [202]. Although not every feature actually contributes equally to privacy preservation and HAR accuracy, we are not concerned with the importance of a single feature, but rather the combination of features that satisfy our goals. Moreover, the solutions are selected by the NSGA-III algorithm.

In contrast to previous feature selection schemes such as [202], we are also concerned about the feasibility of deployment in resource-constrained environments, apart from the most important privacy protection aspect. In this context, resource constraints are mainly

Table 3.3 – Comparison of Three-Objective and Four-Objective Feature Selection Results.

	Classifier	NB	DT	LR	SVM	KNN	XGB	RF
all F	**HAR**	66.58%	81.92%	85.15%	95.21%	96.17%	93.03%	94.39%
	User	21.18%	29.62%	75.56%	93.48%	93.47%	92.50%	94.88%
3 objs	**HAR**	56.74%	68.39%	64.49%	76.28%	74.16%	76.90%	81.07%
	User	7.10%	22.88%	10.55%	34.13%	40.29%	48.74%	56.37%
4 objs	**HAR**	64.91%	77.34%	66.79%	77.25%	77.87%	81.28%	**84.34**%
	User	4.77%	16.76%	4.49%	22.75%	30.43%	39.25%	**45.32**%

associated with the limited processing, memory, and energy capacities of diverse IoT devices. However, under our envisioned architecture, an additional consideration arises from the computational capabilities of edge servers, which form an intermediate layer between local devices and the HAR cloud. Although the above results have demonstrated that the four-objective method is better, in order to compare them more thoroughly, we modify the number of generation of the NSGA-III and observe how this would affect the results of the two classifiers M_1 and M_2 (see Figure 3.6). The number of generations varies: 10, 20, 30, 40, and 50. In Figure 3.6, at generation 0, it corresponds to the results of recognizing activities and users using all 129 features. As generations progress, until the generation reaches 30, the results of the two approaches show little difference, but the four-objective scheme performs slightly better than the three-objective scheme. Subsequently, however, the gap between the two schemes quickly widens, and when the generation is 50, the four-objective scheme has obtained results that satisfy our goal, while the user recognition rate of the three-objective scheme remains high above 50%. Therefore, our proposed method can greatly reduce the number of generations and speed up the convergence of the NSGA-III. This not only significantly reduces computation time on the edge server, but also lowers the resource requirements of devices and accelerates on-device processing, as fewer features need to be extracted, enabling it to be deployed in resource-constrained environments.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

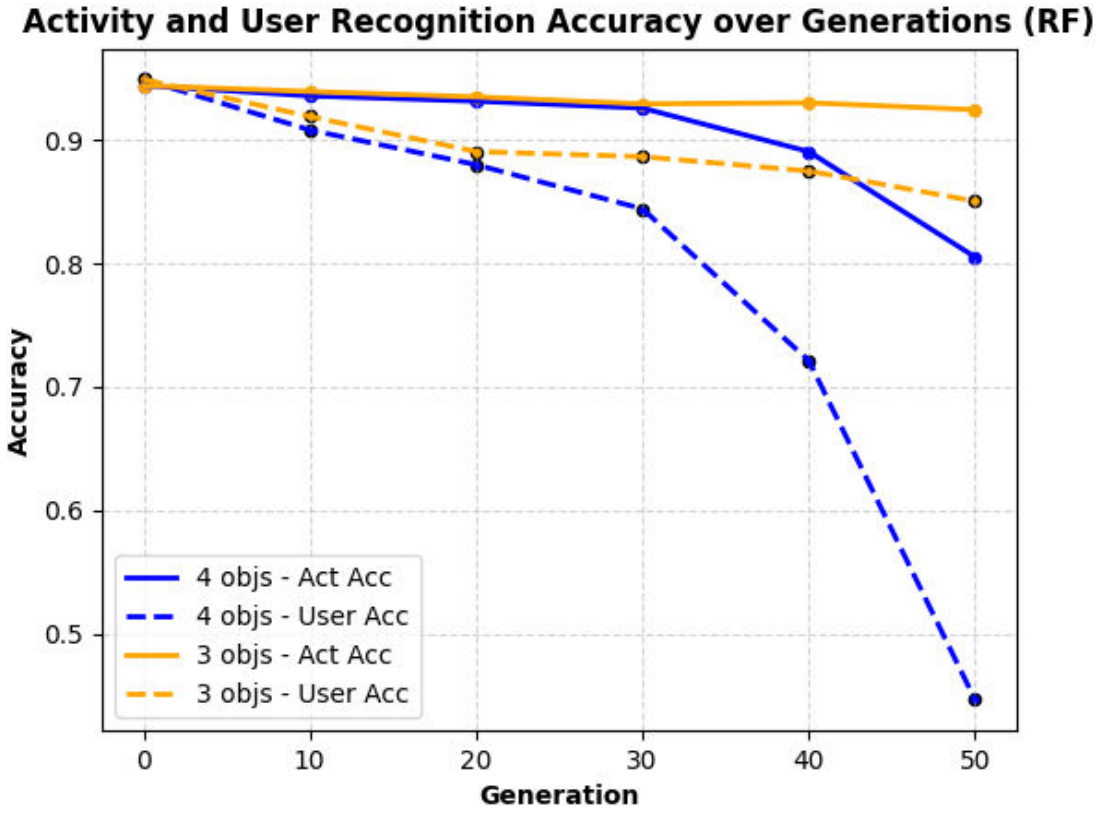


Figure 3.6 – Comparison of HAR and User recognition results using features selected by NSGA-III with four objectives and three objectives.

3.3.2 Comparison of ways to select the optimal features

In each generation, 455 individuals are generated, i.e., 455 different solutions. The key challenge is how to select the desired solution from these populations, because among these solutions, there does not exist an optimal solution that is best on all objectives, but only Pareto optimal solutions. Is it preferable to search for ideal solutions that balance all objectives simultaneously (i.e., all O_1, O_2, O_3 and O_4), or to focus on solutions that emphasize one specific objective of higher importance? Maximizing the difference between HAR and user identification (i.e., O_3) can reasonably seem to be the most important task that is why it was privileged before. In this section, we compare the effectiveness of the feature subsets selected based on the balance of four objectives with those selected based on O_3 in order to validate that hypothesis. In this experiment, the fitness function is still defined with four objectives. The results of HAR and user recognition corresponding to the balanced selection mechanism are presented in Figure 3.7.

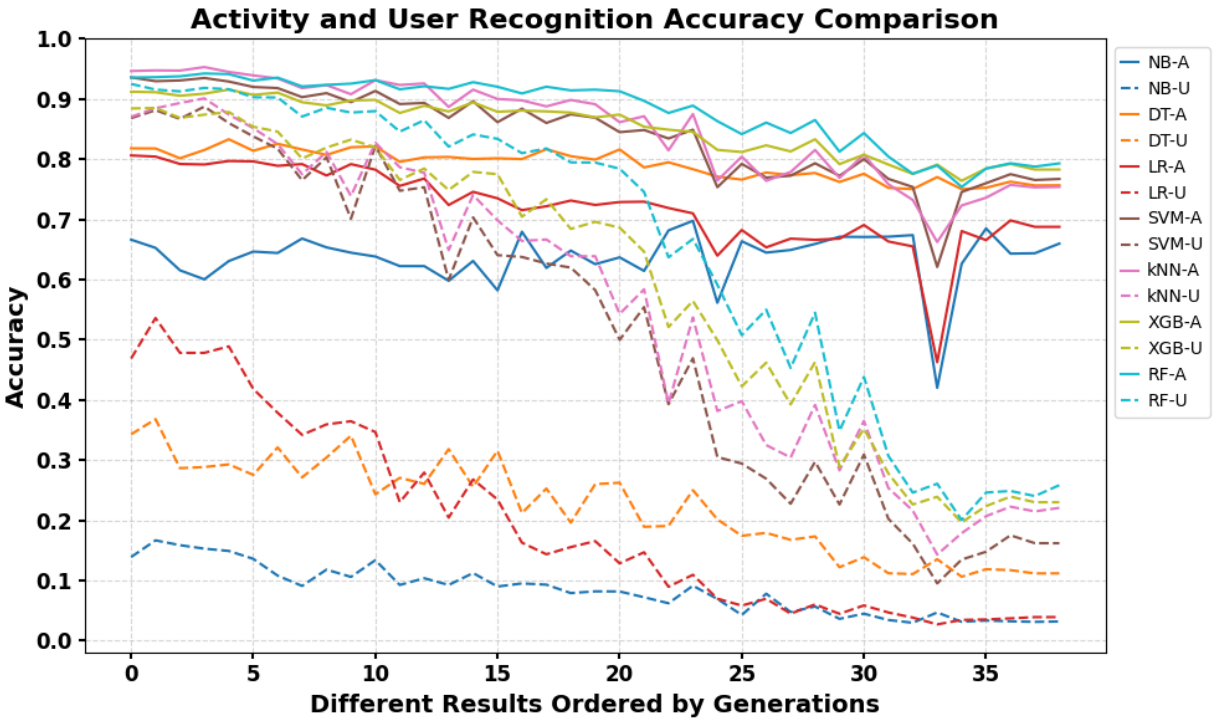


Figure 3.7 – Balanced Solutions (129 initial features): HAR and User identification results using features selected by NSGA-III with four objectives (A indicates activity recognition accuracy, U indicates user recognition accuracy).

The results in Figure 3.5 are obtained by selecting the optimal O_3 in each generation. In general, the accuracies of the HAR and user recognition fluctuate more for the O_3-based scheme (Figure 3.5) compared to the balance-based scheme (Figure 3.7). This may be due to the greater stability of the final solution selection mechanism in the balanced approach. Compared to the strategy that selects the final solution solely based on a single objective considered the most important, the balance-based scheme takes all four objectives into account when evaluating the generated population. This leads to a more equitable trade-off among the multiple objectives, potentially reducing solution bias and extremity, and thereby resulting in greater overall stability in the outcomes. In addition, the HAR accuracy of the balanced approach eventually stabilizes at around 80%, and the user recognition accuracy is about 25%. In contrast, the O_3-based approach finally achieves around 65% accuracy in HAR, and the user identification accuracy is approximately 10%. Therefore, from the perspectives of data availability and data privacy, the feature subsets provided by the balanced scheme appear to better align with our goal. However, according

to the optimal solutions of both approaches shown in Table 3.4, the $Mean_d$ of the balance-based scheme is 52.69, and it is 52.29 for the O_3-based scheme, so their results are very similar. Moreover, these two solution selection mechanisms offer some of the same results, thus, the difference between them is not significant. That said, the balanced approach can sometimes provide better solutions, for example, when the HAR accuracy rate is approximately 81%, the O_3-based method achieves a user recognition rate of 42.47%, whereas the balance-based method achieves 34.86%. The selected features corresponding to the balanced solution selection mechanism in Table 3.4 are: *skewness*, *mean absolute difference*, *peak to peak distance*, *zero crossing rate*, *fundamental frequency* of the y-axis and *interquartile range* of the z-axis.

Table 3.4 – Comparison of O_3-based and Balance-based Feature Selection Results.

	Classifier	**NB**	**DT**	**LR**	**SVM**	**KNN**	**XGB**	**RF**
all F	**HAR**	66.58%	81.92%	85.15%	95.21%	96.17%	93.03%	94.39%
	User	21.18%	29.62%	75.56%	93.48%	93.47%	92.50%	94.88%
O_3	**HAR**	64.91%	77.34%	66.79%	77.25%	77.87%	81.28%	84.34%
	User	4.77%	16.76%	4.49%	22.75%	30.43%	39.25%	45.32%
bal.	**HAR**	67.07%	77.57%	69.06%	80.01%	80.75%	80.74%	**84.33%**
	User	4.49%	13.87%	5.86%	30.92%	36.51%	35.21%	**43.82%**

3.3.3 Impact of different initial features

In the previous experiments, a total number of 129 features were selected. Now, we want to explore the impact of adding more initial features. These three features, *Wavelet absolute mean*, *Wavelet energy* and *MFCC*, are added to the original feature set. These were chosen because wavelet-based features capture signal magnitude and energy in the time–frequency domain, while MFCC provides a compact representation of the spectral envelope, thus complementing the existing feature set. Each of them has multiple values, and considering that the accelerometer includes x, y and z axes, the total number of features increases by 90. Therefore, the total number of features is now 219. In order to obtain satisfactory results, here the number of generations is increased to 150, the final solutions in each generation are selected in a balanced manner (the four objectives are

used), and other parameters are kept the same. The results are shown in Figure 3.8.

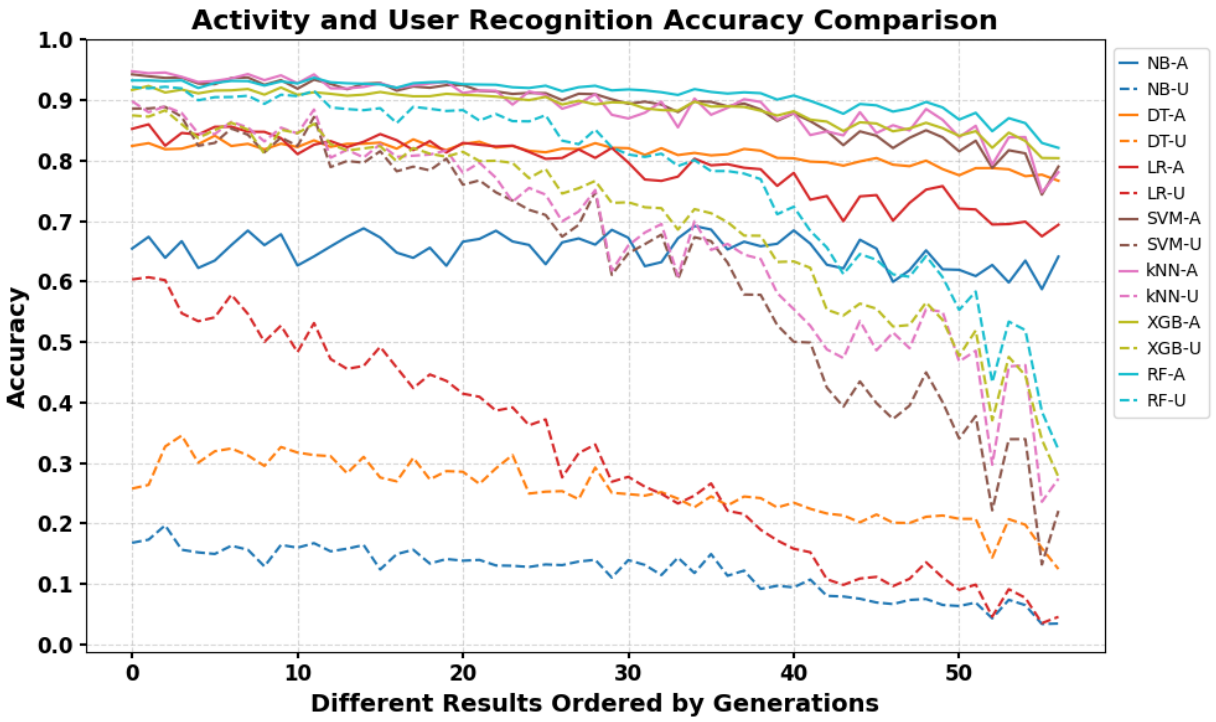


Figure 3.8 – Balanced Solutions (219 initial features): HAR and User identification results using features selected by NSGA-III with four objectives (A indicates activity recognition accuracy, U indicates user recognition accuracy).

Compared to the use of 129 initial features (see Figure 3.7), employing 219 initial features does not lead to a significant change in user identification accuracy during the early stages (see Figure 3.8). This may be because, although more features provide richer information and a larger solution space, the NSGA-III algorithm has not yet effectively utilized this information at the beginning of the search, resulting in its advantages not yet being apparent. In addition, the increased feature dimensionality may introduce more redundant or less relevant features, which in turn increases the complexity of the search, thereby limiting the improvement in algorithm performance in the early phase. A noticeable downward trend only begins to appear around the 55th generation (i.e., the 38th solution). Furthermore, the final user recognition accuracy achieved using 219 initial features is slightly higher than that obtained using the 129 initial features. However, by appropriately increasing the number of generations, we are still able to find the desired solution, as shown in Table 3.5. The $Mean_d$ of the optimal solution with 129 initial fea-

tures is 52.69, whereas that with 219 initial features is 54.35 which is slightly better (1.66 points). The corresponding selected features are: *fundamental frequency* of the x-axis, *variation coefficient*, *mean absolute difference*, *sum of absolute difference*, *peak to peak distance*, *zero crossing rate*, *fundamental frequency* and *mean frequency* of the y-axis.

Table 3.5 – Comparison of results for 129 initial features and 219 initial features.

	Classifier	**NB**	**DT**	**LR**	**SVM**	**KNN**	**XGB**	**RF**
all 129 F	**HAR**	66.58%	81.92%	85.15%	95.21%	96.17%	93.03%	94.39%
	User	21.18%	29.62%	75.56%	93.48%	93.47%	92.50%	94.88%
subset 129 F	**HAR**	67.07%	77.57%	69.06%	80.01%	80.75%	80.74%	**84.33%**
	User	4.49%	13.87%	5.86%	30.92%	36.51%	35.21%	**43.82%**
all 219 F	**HAR**	66.48%	83.43%	89.41%	95.52%	95.93%	93.50%	94.11%
	User	21.48%	29.27%	81.58%	94.04%	93.58%	91.61%	94.91%
subset 219 F	**HAR**	62.75%	78.77%	69.44%	78.79%	79.34%	82.12%	**84.87%**
	User	4.33%	14.38%	4.61%	22.18%	29.73%	37.06%	**43.36%**

3.4 Conclusion

This chapter addresses the dual challenges of privacy protection and resource-constrained deployment in assisted living scenarios by proposing a feature selection algorithm that balances computational efficiency and privacy protection. The approach is designed with full consideration of the computational and storage constraints of the target system. By introducing a convergence acceleration mechanism (i.e., O_4), the privacy-preserving feature selection framework can more quickly identify the optimal or near-optimal combination of features, thereby effectively reducing computational overhead and improving response speed. Experimental results demonstrate that, compared to previous work, the algorithm proposed in this chapter not only maintains competitive privacy protection performance but also significantly reduces the time (or the number of generation) required for feature selection, making it particularly suitable for assisted living systems with limited computational and storage resources.

It is worth noting that the research in this chapter still has certain limitations. First, the performance verification of this algorithm is based solely on a single attacker model (the supervised learning attack introduced in Chapter 2), so its privacy protection capabilities may decline in more complex or diverse attack scenarios. In addition, although such attacks are the most severe in terms of threat level, their characteristics often make them a priority target for prevention in practical applications. Therefore, since defensive measures are deployed at the initial stage of the HAR system, the probability that such attacks occur in real-world environments is relatively low. However, our privacy-preserving feature selection method can still ensure the privacy of individuals with disabilities even after data leakage, as these features are insufficient for AI models to learn the connection between data and identity. Second, the deployment and operation of our algorithm in real-world assisted living environments have not yet been tested. Although the results from experimental environments suggest feasibility, further engineering validation is required, including assessments of stability and resource consumption under different hardware platforms, network conditions, and long-term operational states.

In next chapter, we will focus on expanding and integrating the testing and analysis of multiple attacker models (detailed in Chapter 2) to enhance the robustness and applicability of the algorithm in multi-threat environments. Through this effort, it is expected that the approach proposed in this chapter will mature further, becoming an important candidate solution for privacy-preserving feature selection techniques in resource-constrained

scenarios, thereby providing a robust foundation for the secure and efficient operation of smart assisted living systems.

Chapter 4

Robust Privacy-preserving Feature Selection in Human Activity Recognition Systems

As assisted living systems become increasingly widespread, daily activities, health status, and interaction data of people with disabilities are continuously collected and processed. The question of how to provide intelligent services while ensuring user privacy has become a core issue that cannot be ignored in the design and deployment of such systems. In the previous chapter, we proposed a privacy-preserving feature selection algorithm designed for resource-constrained environments, and experimentally validated its effectiveness under a single attacker model (i.e., supervised learning attacks). However, real-world attack threats are far more complex and diverse than this single-attacker model. Attackers vary significantly in terms of their background knowledge, attack strategies, and resource capabilities. Since the previous chapter focused primarily on algorithm design and validation in a single-attacker scenario, its coverage and analysis of diverse attacker models remain inadequate, limiting the robustness and security assessment of the algorithm in practical applications.

From the perspective of data privacy, robustness depends not only on the performance of the algorithm performs under a single attack, but also on its adaptability in environments where multiple threats coexist (i.e., different attacker models) and attack patterns (i.e., different AI inference models) are uncertain. Privacy threats to assisted living systems may originate from malicious operations by internal legitimate participants, or from passive eavesdropping or active inference attacks by external parties. These attacks show significant differences in terms of their targets, resource consumption, and methods of information exploitation. Only approaches that can maintain privacy protection performance across a wide range of attack scenarios can be considered to possess practical robustness. Therefore, conducting systematic testing of algorithm performance

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

under a multi-attacker model is a necessary step to verify whether it can operate reliably in real-world deployments over the long term.

Given this reality, this chapter will focus on analyzing three attacker models. These three models encompass the supervised learning attack scenarios already discussed in the experiments of the previous chapter, as well as two additional attacker models that were described in detail at the conceptual level in Chapter 2: semi-supervised learning attacks and unsupervised learning attacks. All these models are based on threat assumptions in cloud environments and reflect the different types of attack scenario that assisted living systems may face in the cloud. The discussion in Chapter 2 focused on clarifying the threat assumptions and potential scenarios of these attacker models, while the objective of this chapter is to introduce them into the experimental validation phase, i.e., using real-world datasets, designing and implementing corresponding attack experiments to quantify and demonstrate the threats posed by different attackers to user privacy, thereby systematically evaluating the robustness of our algorithms in a multi-model threat environment. On the other hand, multi-model threats also manifest in the possibility that attackers may employ any advanced AI inference model, which exposes potential vulnerabilities in the Random Forest (RF) based approach discussed in the previous chapter when facing a broader family of models. In real-world scenarios, however, the models and inference mechanisms adopted by attackers are often more complex and diverse, which is precisely the problem this chapter aims to address.

As shown in the previous chapter, privacy protection research has proposed a variety of technical approaches, including local differential privacy, data sanitation, homomorphic encryption, and federated learning, etc. For example, local differential privacy provides provable privacy protection for the publication of statistical information [217]; homomorphic encryption is applied mainly to the encrypted sharing of model weights in federated learning, ensuring privacy during collaboration [108]. In addition, some studies have attempted to focus on the robustness of the methods in their design and evaluate their performance under different attack scenarios. However, as pointed out in [218], systematic analysis of the relationship between privacy and robustness in federated learning is still insufficient, and its coverage of diverse threat models (such as the roles and objectives of different attackers) is still incomplete. Furthermore, although some studies have proposed robust methods for privacy protection, these methods are difficult to independently verify and reproduce due to their overly complex designs and lack of flexible open-source implementations. This lack of verifiability in the experimental section makes it difficult to

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

be convinced of the claimed robustness guarantees [219]. These current conditions reflect that, in the field of privacy protection, especially in resource-constrained scenarios (including various local devices and edge servers, etc.), introducing a multi-attacker model for systematic verification has both practical significance and significant research value.

In this chapter, we will first use the HASC-IPSC dataset to conduct experiments on an unsupervised learning attacker model to verify the extent of its threat to user privacy (i.e., the user identity). This process not only quantifies the actual impact of various attacks, but also provides a basis of attack scenarios for the robustness testing of our algorithm in subsequent sections. The definition and assumption of the specific attacker model have been conceptually described in Chapter 2. This chapter will not elaborate on them, but will instead focus on validation and analysis based on experimental data. Through this arrangement, threat modeling can be more directly linked to experimental evaluation, making robustness analysis more empirical and targeted.

4.1 Threat Model Instantiation and Analysis

This section focuses on the instantiation of the threat model introduced in Chapter 2, by showing how the considered attacks can be practically implemented. It then examines whether the protection mechanism proposed in Chapter 3 is sufficient to withstand the instantiated threat. The experimental verification of the attack model is based entirely on the HASC-IPSC dataset. We assume that the attackers are located on the server side, i.e., it is possible for the attackers to access the user data uploaded to the cloud, which in our case is the subset of features and activity labels of the user's sensor data. Additionally, it is important to emphasize that the activity label is considered public information. In this context, "user privacy" refers to the user's personal information such as identity, gender, height, etc.

The supervised learning attacks were examined in Chapter 3, while in this chapter we make the assumption that the proposed method can also defend against semi-supervised learning attacks. Therefore, unsupervised learning attacks will be our main focus in this section. The difference between these two is that the former attackers may have background knowledge such as "some" user identity or biometric information, and are therefore assumed to pose a greater threat to user privacy, whereas the latter attackers only have access to the data transmitted to the cloud and no background knowledge.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

4.1.1 Unsupervised Learning Attacks

In Chapter 3, we assumed that the retention of features that represent activities rather than identities is sufficient to avoid user recognition. However, attackers may use unsupervised learning models to attempt to cluster data from different users as a way to perform user privacy inference attacks. As an illustration of such an attack, Figure 4.1 shows the purities of the clusters obtained by applying K-means to the optimal feature subset selected according to the method in Chapter 3. We set the purity threshold to 0.7 to determine whether a cluster is pure, i.e., when more than 70% of the samples in the cluster are from the same category (i.e. identity in our case), the cluster is considered pure, otherwise it is considered impure. This choice is based on two considerations. First, to ensure the feasibility of subsequent analysis, we required at least $M_{min} = 8$ clusters (for the full dataset) to satisfy the purity threshold, and 0.7 represents the highest threshold meeting this minimum cluster count (higher thresholds would result in too few analyzable clusters). Second, under the null hypothesis of random label assignment, the expected number of false positive clusters at $t = 0.7$ is extremely low, indicating that clusters reaching this purity are highly unlikely to occur by chance. Finally, for clusters labeled as pure, we further computed their entropy and normalized entropy, which showed that these clusters are highly concentrated internally (entropy values near zero or well below the expected values under random distribution), providing additional evidence of their reliability. Taking into account the number of analyzable clusters, statistical significance, and internal concentration, the threshold of 0.7 was selected as a reasonable compromise, though it can be adjusted if different trade-offs are desired.

From the figure, it can be seen that even without the user label information, the attacker is still able to identify two users with the best subset of features. Moreover, attackers could launch multiple attacks for cross-inference, thereby inferring the label of additional users. Therefore, the approach presented in Chapter 3 is not capable of defending against such attacks, as attackers can exploit the incompleteness of the assumption to identify a few individuals using the selected feature subset. It is therefore important to propose a new robust framework that can protect user privacy against these attackers as well as the more powerful one, as the latter still may be able to obtain user labels for some of the data through other means, and thus can distinguish data from different users with the help of stronger models. As demonstrated above, existing work cannot address our problem after selecting the best subset of features using a supervised model. Therefore, next section will propose our robust approach that can defend against the above attack

and that does not compromise the performance of existing work.

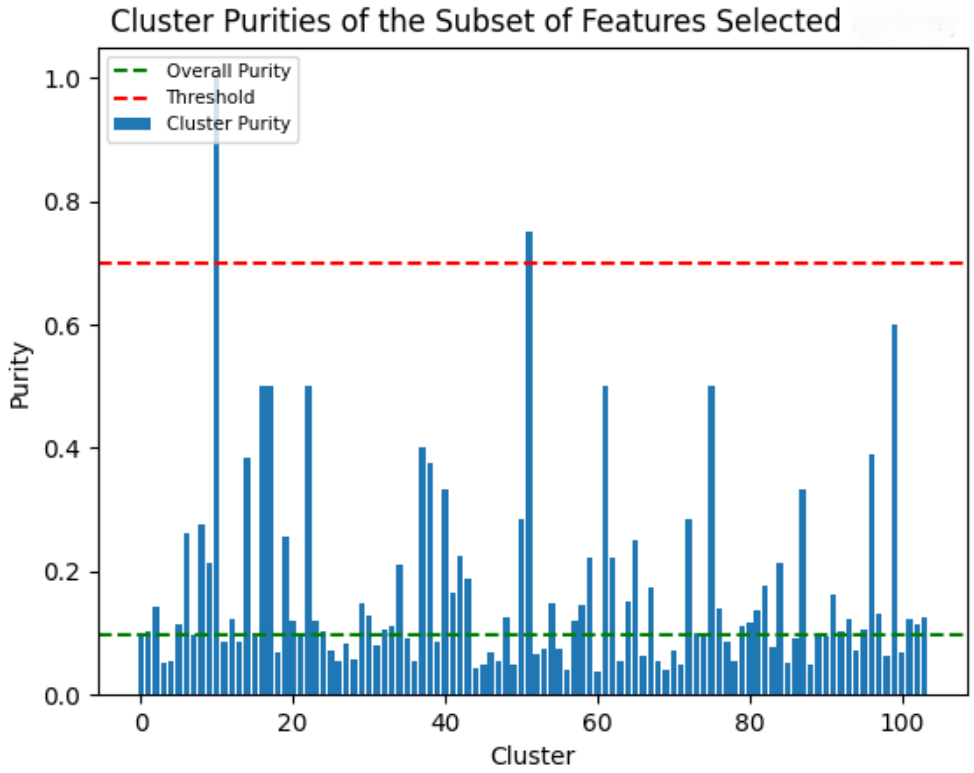


Figure 4.1 – Purity of Each Cluster Using the Subset of Features Selected According to the Approach in Chapter 3 (HASC-IPSC dataset).

4.2 Robust Privacy-Preserving Feature Selection

In this section, our robust model is presented for selecting privacy-preserving features in sensor-based HAR systems, as shown in Figure 4.2. Thus, the complete objective definition of our feature selection algorithm will be stated, and then the robustness of our proposed model will be explained.

4.2.1 Robust Privacy-Preserving Feature Selection

Our goal is to improve the robustness of privacy-preserving feature selection against user privacy inference attacks by ordinary attackers (i.e., unsupervised learning attacks), as well as other possible inference attacks by powerful attackers (i.e., semi-supervised

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

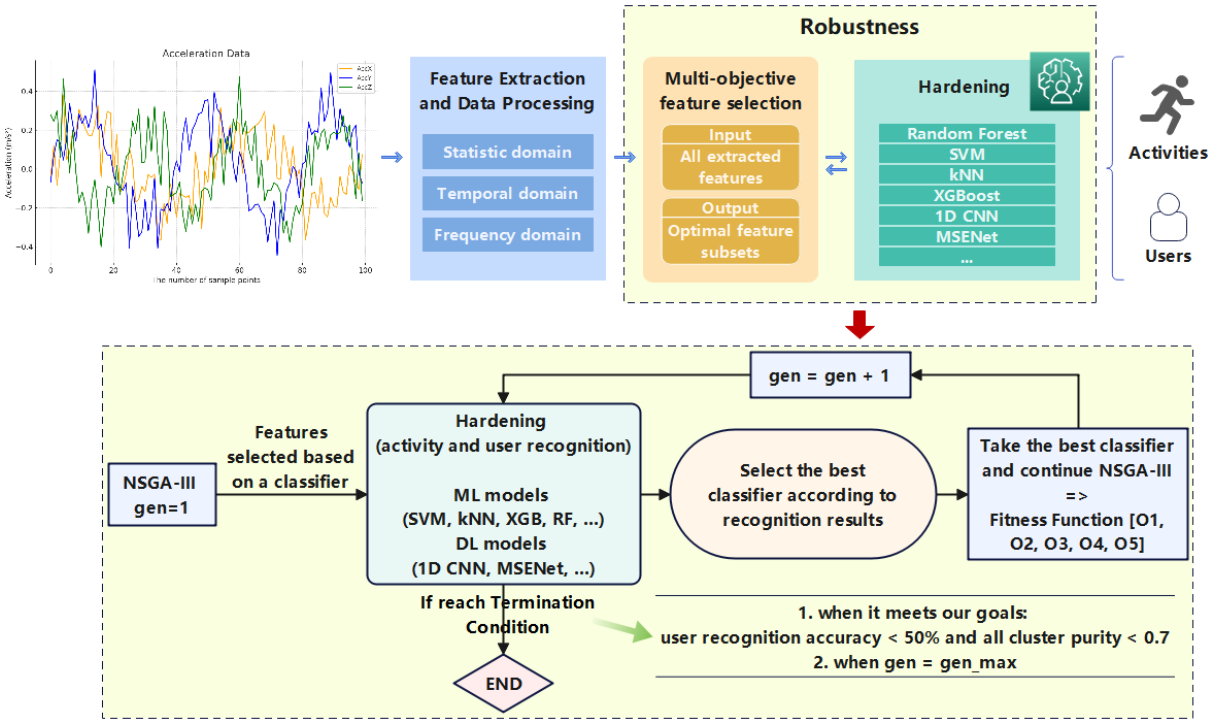


Figure 4.2 – The robust framework for privacy-preserving feature selection.

and supervised learning attacks). Additionally, it is also necessary to maintain a high activity recognition accuracy. This task is also performed with the help of Multi-Objective Evolutionary Algorithms (MOEA), more precisely the Non-dominated Sorting Genetic Algorithm III (NSGA-III). For a detailed introduction and explanation of MOEA and NSGA-III, please refer to Chapter 3.

To achieve our goal, we follow the objective definitions in the previous chapter (see 3.1.1) for the selection of the optimal subset of privacy-preserving features: O_1, O_2 and O_3 aim to select features that represent activities rather than identities, and the goal of O_4 is to speed up the convergence of the NSGA-III algorithm by reducing the number of selected features. It is assumed here that O_2 and O_3, in addition to being able to help resisting against supervised learning attacks, are also capable of defending against semi-supervised learning attacks; therefore, no additional objectives are set specifically for semi-supervised learning attacks. Moreover, to defend against user privacy inference attacks by ordinary attackers, a new objective is added to our framework in order to enhance its robustness in privacy protection compared to the original: lower the purity

of all clusters to prevent user re-identification. To this end, K-means is utilized to cluster the data (i.e., the subset of features), and then the purity of each cluster is computed. Next, the number of normalized clusters with purity less than 0.7 is calculated as the objective value. Therefore, the new objective of the NSGA-III algorithm is defined as follows. For better explanation, all clusters are defined as C (its number is equal to the total number of users) and c represents the number of clusters with purity less than 0.7. The fitness function with complete objective definitions is specified in Algorithm 2, which aims to evaluate the quality of the solutions and provide optimization directions for the populations of the NSGA-III algorithm by computing the objective values.

— The fifth objective O_5 aims to maximize the number of clusters c to avoid user inference attacks:

$$O_5 = max(\frac{|c|}{|C|}) \tag{4.1}$$

Algorithm 2 Fitness Function

1: **procedure** FITNESS FUNCTION
2: Train the HAR model C_1
3: Train the User identification model C_2
4: Calculate the accuracy of C_1: O_1 (see [220])
5: Calculate the accuracy of C_2: O_2 (see [220])
6: Calculate the performance difference between C_1 and C_2: O_3 (see [220])
7: Calculate the normalized number of selected features: O_4 (see [220])
8: Calculate the normalized number of clusters with purity less than 0.7: O_5 (see Eq. 1)
9: **return** $-O_1, O_2, -O_3, O_4, -O_5$ (*minimize* function is used in NSGA-III)
10: **end procedure**

Building on the previous chapter, the purpose of this chapter is to propose a robust privacy-preserving model for selecting a subset of features that achieve high activity recognition accuracy and protection of user privacy in a local environment. O_5 could defend against user privacy inference attacks by ordinary attackers. Whereas O_1, O_2 and O_3 are used to select features indicative of activities instead of identities with a supervised learning model (these objectives are also assumed to be effective in protecting user identity against semi-supervised learning attacks), but this may not be sufficient. In fact, RF was utilized in the previous chapter as the only classifier for the NSGA-III algorithm. However, malicious attackers may employ more powerful supervised learning models, so to protect the user from the most powerful adversaries, it should be assumed that the

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

best classifier is always used in the NSGA-III algorithm in order to further strengthen the robustness of our approach. Consequently, during each generation of the NSGA-III algorithm, our proposed model selects the classifier that performs the best among a list of possible classifiers employed by the attacker as the classifier for selecting the feature subset in the next generation of the NSGA-III algorithm, i.e., after the feature selection is completed in each generation, our proposed model verifies the selected feature subset with all possible classifiers on both activity and user recognition tasks and then selects the one with the best performance for the next generation. To determine the optimal classifier, we compute the product of the accuracies from the two classification models and select the result with the highest value, as we consider a classifier to be optimal only if it performs well on both tasks. The process of our proposed model is shown in Algorithm 3. In this way, we can ensure that the subset of features selected in each generation is robust in terms of privacy preservation, further improving the level of user privacy protection.

Therefore, Algorithm 2 and Algorithm 3 achieve together the robustness of our privacy-preserving solution, as shown in Figure 4.2. The method initially selects a classifier at random. Subsequently, after the generation ends, the best classifier is selected according to Algorithm 3. Then, new populations are created and evolved based on Algorithm 2 (i.e., the NSGA-III algorithm). The repetition of Algorithm 2 and Algorithm 3 continues in this manner until the termination condition of the approach is reached: 1) considering the resource constraints of the environment, when the user identification of the best classifier is below 50% (i.e., it meets privacy protection requirements by preventing accurate user recognition), and the number of normalized clusters with purity less than 0.7 is 100% (i.e., the purity of all clusters is below 0.7), the algorithm is stopped; 2) when the NSGA-III algorithm reaches the maximum number of generations. Two termination conditions are designed, and the specific choice between them depends entirely on the local environment (i.e., the capability of the edge server). A detailed comparison of them will be discussed in the experimental results section.

4.3 Experimental Evaluation

This section presents our experimental results. To verify our proposed robust privacy-preserving feature selection framework, this chapter uses two datasets: 1) HASC-IPSC and 2) WISDM. Their detailed descriptions can be found in Chapter 2. First, the feature extraction process and the initial features associated with the changes in the time-series

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

Algorithm 3 Robust Privacy-Preserving Feature Selection

Input:
F: all extracted features from acceleration data
$Classifiers$: a list of possible classifiers used by the attacker
$Init_C$: kNN (random selection)
MAX_GEN: maximum generation of the NSGA-III algorithm
TC: Termination Condition 1 or Termination Condition 2
Output:
f: the selected subset of features

Steps:

$f \leftarrow F$
$classifier \leftarrow Init_C$
$gen \leftarrow 1$
$max_gen \leftarrow MAX_GEN$
while *not* TC **do**
    $f \leftarrow$ NSGA-III$(f, classifier)$
    $P \leftarrow []$ ▷ P is the product of the accuracies of the two classification task models
    for $C \in Classifiers$ **do**
        Train the HAR model C_1 with f
        Train the User identification model C_2 with f
        Calculate P of the results of C_1 and C_2
    end for
    $max_index \leftarrow$ MaxIndex(P) ▷ Get the index of the largest value in P
    $classifier \leftarrow Classifiers[max_index]$ ▷ Get the best classifier of this generation
end while
return f

data will be briefly described. Next, all the classification models used in the robust privacy-preserving feature selection framework will be presented. Finally, the experimental setup will be discussed and the experimental results will be analyzed. It should be noted that the experimental results are first obtained on the basis of the HASC-IPSC dataset and then validated with the WISDM dataset.

4.3.1 Feature Extraction

In the experimental verification of this chapter, the initial features extracted are basically consistent with those listed in Table 3.2, and multiple features were obtained from the statistical, temporal, and frequency domain, respectively. Note that in addition to the correlation features used to evaluate the relationship between the different axes of the signal, the remaining 44 features listed in Table 3.2 are extracted separately from the x, y and z axes of the acceleration signal. Furthermore, according to the experimental results in Chapter 3 (see Section 3.3.3), *mfcc, wavelet absolute mean, wavelet energy* are added to the initial list of features, and these three features have multiple values. Therefore, the total number of initial features extracted is $71 \times 3 + 6 = 219$. The extraction of features is facilitated by utilizing the existing Python libraries Tsfel [214] and NumPy [215].

4.3.2 Classifiers

In previous work [203, 208, 202, 220], researchers have chosen RF as the classifier in the NSGA-II or NSGA-III algorithms due to its power, ease of interpretation, and reasonable computational cost [202]. However, in this work, after completing feature selection, we will assess the effectiveness of the subset of features selected by the NSGA-III algorithm using Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), RF, XGBoost (XGB), CatBoost, 1D CNN and MSENet [201]. In other words, these ML or DL algorithms will be utilized to recognize human activities and users, respectively, in order to confirm whether the selected subset of features can protect user privacy while accurately identifying activities. The detailed parameters of these classification models will be provided in the next subsection.

4.3.3 Experimental Setup

All experiments are performed on a Linux machine with 200GB of RAM, a 64 core Intel Xeon Gold 5218R CPU and a NVIDIA RTX A6000 (48Go) GPU. All models are

executed in a Python 3.10 environment. The versions of the main packages used are as follows: *Pymoo* 0.6.1.3, *Numpy* 1.26.4, *Scikit-learn* 1.5.2, *Tensorflow* 2.18.0, and *Tsfel* 0.1.6.

In our experiments, a total of 219 features are extracted from each time window of the dataset to form an initial feature set. Meanwhile, all data are normalized before feature selection. This step ensures that the features have a common scale before they are used in the classification task. Next, these data are randomly divided into two sets: 70% of the data are used as training data (with 20% of them allocated as the validation set) and 30% of the data are reserved as testing data. The NSGA-III algorithm selects the optimal subset of features based on the performance of the features on the validation set. The final evaluation is performed on the test set.

The NSGA-III algorithm is characterized by the following key parameters:

— The population size is 500. The reference point division parameter was set to $p = 8$ to limit the number of reference points and keep the computational cost manageable for the five-objective problem ($M = 5$) in our experimental environment, yielding $H = \binom{p+M-1}{M-1} = 495$ reference points. The number of populations was set to 500, which preliminary tests showed was sufficient for convergence and Pareto front diversity.

— The number of generations is 110. Similar to the four-objective experiments, this choice balances convergence and solution diversity. The slightly lower value compared to 120 in the previous chapter reflects the adjusted experimental goals and the reduced reference point resolution ($p = 8$), which allows sufficient exploration with fewer generations.

— The initial population is generated using BinaryRandomSampling.

— Simulated binary crossover with default parameters [207] is used to ensure efficient exploration by combining parent solutions.

— The polynomial mutation operator [207], using a probability of 0.1 and an eta of 20, introduces controlled variation to avoid premature convergence.

— Duplicate solutions are removed to maintain population diversity throughout evolution.

As for the parameters of each classifier: SVM uses the RBF kernel, with C = 10 and gamma = *'scale'*; and the selected parameters for CatBoost are *iterations* = 500, *learning_rate* = 0.1, *depth* = 4, *loss_function* = *'MultiClass'*. Please refer to [216] for

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

the parameters of other ML classification models. The 1D CNN model has two layers, the first convolution layer has 64 feature maps and a kernel size of 3 and the second convolution layer has 128 feature maps and a kernel size of 3. Then, they are followed by a dropout layer for regularization and a max pooling layer. After that, the learned features are flattened into a long vector, and then they are fed into a dense layer before passing through the output layer used for activity or user identification. In addition, the structure of the MSENet is described in detail in [201], and in our work, most of the parameters remain the same as in [201], except that 64 feature maps are used in the first and third convolution layers and 128 feature maps are used in the second convolution layer, as well as a dropout layer is added after the global average pooling layer. Furthermore, the number of clusters for K-Means is fixed at 105, which is consistent with the number of users.

4.4 Experimental Results

Our goal is to propose a robust privacy-preserving approach that can be deployed in a resource-constrained environment (including both edge computing and local devices) to address various potential privacy threats during data transmission and application. Compared to the previous chapter, this chapter is more interested in the robustness of our approach, i.e. it needs to be able to defend against many possible attacks. In the following subsections, we will separately analyze the robustness of our framework on the threat models discussed in Chapter 2.

4.4.1 Robustness Testing under Supervised Learning Attacks

First, let us discuss the robustness of our approach against powerful supervised learning models. This is related to the objectives O_1, O_2 and O_3 defined in our NSGA-III algorithm: balancing user privacy protection with maintained relatively high HAR accuracy. All the supervised learning models mentioned above are used to simulate possible privacy attacks, such as k-NN, RF, MSENet. Furthermore, it should be noted that in each generation of the NSGA-III algorithm, the optimal solution is chosen based on the key objective O_3, which aims to maximize the divergence between the two classification tasks. i.e., C_1 and C_2. Figure 4.3 shows the accuracy trends across generations for activity and user identification using features selected by the NSGA-III algorithm. To ensure clarity, only

distinct results in generational order are retained, so the final generation number shown in the figure is not the maximal generation.

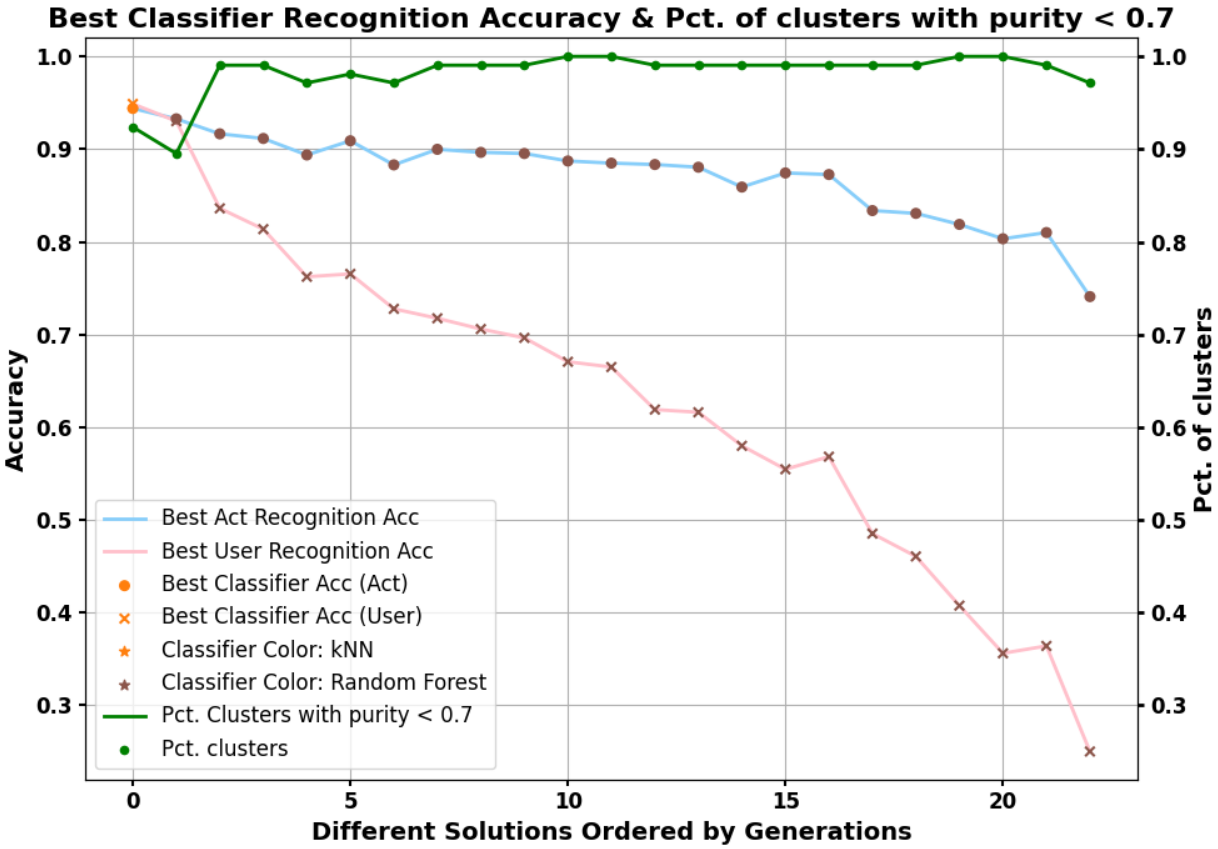


Figure 4.3 – Robust NSGA-III Feature Selection Effects on HAR Accuracy, User Recognition Accuracy, and Cluster Purity of HASC-IPSC dataset.

In Figure 4.3, the left axis represents the accuracy of activity or user recognition, while the right axis represents the percentage of clusters with purity less than 0.7. The green curve corresponds to the right axis, whereas the pink and blue curves are linked to the left axis. At generation 0, the HAR and user identification accuracies, as well as the percentage of clusters with purity less than 0.7, are obtained using all 219 features. In addition, kNN is randomly selected as the initial best classifier, so that at generation 0, the color of the marked points on the curve is orange. However, it can be seen from the figure that the best classifier switched back to RF from the first generation until the end of the feature selection algorithm, as all data points are marked in brown. Therefore, the approach presented in the previous chapter proved to be relatively robust from this perspective.

However, since the contribution of this chapter lies in proposing an evolutionary and more robust framework, it is necessary to further verify its adaptability and effectiveness under more diverse conditions.

As shown in Figure 4.3, starting from the 17th solution, our initial goal has already been achieved for user recognition accuracy (i.e., falling below 50%), but it has not yet achieved the purity requirement. Considering the resource constraints of the environment and the need to ensure robust privacy protection, two termination conditions of the NSGA-III algorithm have been designed in Section 4.2. Therefore, for Condition 1, the optimal solution corresponds to the solution obtained at the final generation of the algorithm, while for Condition 2, the NSGA-III algorithm stops when the maximum number of generations is reached, and then the optimal solution is selected from all possible solutions that satisfy Condition 1, i.e., the solution space is defined only for cases where 100% of the clusters have a purity below 0.7 (with purity 100%). To evaluate different solutions quantitatively, a scoring metric (score) is designed and the solution with the highest score is selected as optimal. The score is computed as:

$$score = \alpha * norm_{diff} + (1 - \alpha) * norm_{act} \tag{4.2}$$

where $norm_{diff}$ denotes the normalized difference between the two classification models of the best classifier and $norm_{act}$ denotes the normalized HAR accuracy of the best classifier. α weights the relative contributions of the two components, and our aim is to maximize the difference while selecting the solution with higher HAR accuracy, so here α = 0.7.

According to Figure 4.3, the best solution obtained according to Condition 1 is the 19th solution and the optimal solution selected based on Condition 2 is the 20th solution. However, both solutions can robustly protect the user's privacy, as the user recognition accuracy is only about 40% even for the 19th solution, and they both ensure that the purity of each cluster remains below 0.7. The comparison of these two solutions is shown in Table 4.1. Since we focus on the protection of users' private information, the solution obtained from Condition 2 is our optimal choice. Compared with the best-performing result that uses raw data to recognize activities and users, the user recognition accuracy of the optimal solution drops by 60.04%, while the HAR accuracy loses only 17.73%. The corresponding selected subset of features are the following: *range*, *area under the curve*, *zero crossing rate*, *fundamental frequency*, *wavelet energy 2* of the y-axis and *mean difference* of the z-axis.

Table 4.1 – Comparison of recognition results using: Raw Data vs 219 Features vs Subset of Features selected by NSGA-III.

Data Type	Row (MSENet)	219F (kNN)	219F (RF)	subsetF (cond. 1)	subsetF (cond. 2)
HAR	98.06%	95.93%	94.11%	81.88%	80.33%
User	95.65%	93.58%	94.91%	40.75%	35.61%

4.4.2 Robustness Testing under Semi-supervised Learning Attacks

In this subsection, we simulate slightly weaker attackers, i.e., they know fewer user labels. In our experiments, 15% of the training data is randomly selected as the data whose user labels are known to the attackers. In addition, three classifiers, RF, XGB, and Dense Neural Network (DNN), are chosen to validate the robustness of our framework by using all 219 features, as well as the optimal subset of features to recognize activities and users, respectively. The parameters of RF, and XGB are detailed in the previous subsection. The DNN consists of three layers (128, 64, and 32 units, respectively) with ReLU activation. A dropout layer with a dropout rate of 0.2 follows each of the first two layers. The output layer has *num_classes* units with a softmax activation function. The results are shown in Table 4.2.

Table 4.2 – Semi-Supervised Learning Results of HAR and User Recognition using all 219 features and the optimal subset of features (HASC-IPSC dataset).

	Classifier	DNN	RF	XGB
all $219F$	**HAR**	88.87%	86.81%	89.49%
	User	62.99%	72.51%	57.25%
subset F	**HAR**	75.92%	75.53%	75.73%
	User	6.83%	21.14%	15.28%

As can be seen in Table 4.2, if the 219 features extracted from time-series data are transmitted directly to the HAR system server, although the attacker has only a small portion of the labeled data, there is still a risk of privacy leakage because the attacker remains able to identify different users to some extent. However, if only the optimal subset

of features is transmitted, it is difficult for the attacker to distinguish different users, even if the attacker has a portion of labeled data. Therefore, our method can defend against semi-supervised learning attacks.

4.4.3 Robustness Testing under Unsupervised Learning Attacks

Finally, the ordinary attackers are considered. Here, the attacker has access to the features and the associated activity label. *K-means* is used to cluster the data with 219 features and the optimal subset of features, respectively. The purity of each cluster is computed to verify if an attacker can recognize some users using unsupervised learning methods. The results are shown in Figure 4.4 and Figure 4.5.

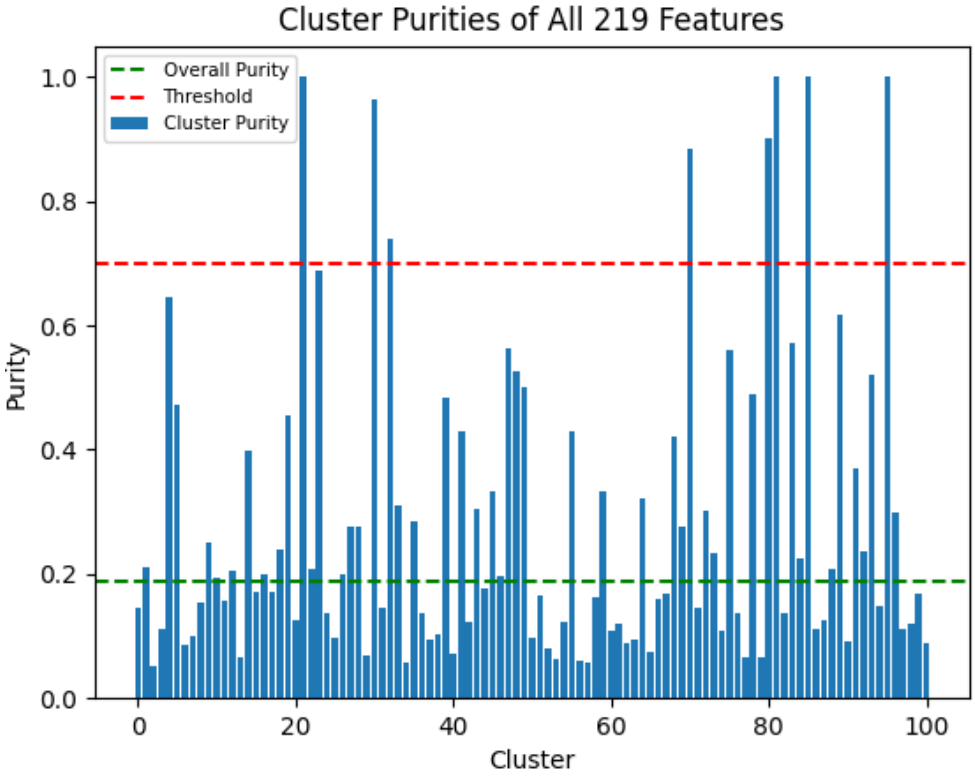


Figure 4.4 – Purity of Each Cluster Using All 219 Features (HASC-IPSC dataset).

As shown in Figure 4.4, even though the average purity of all clusters is less than 20%, in fact, a fraction of users can still be recognized by the attacker. The threshold in the figure is set to 70%, which is consistent with that used in the NSGA-III algorithm. It can be seen that some of the user's data have been effectively clustered together, therefore

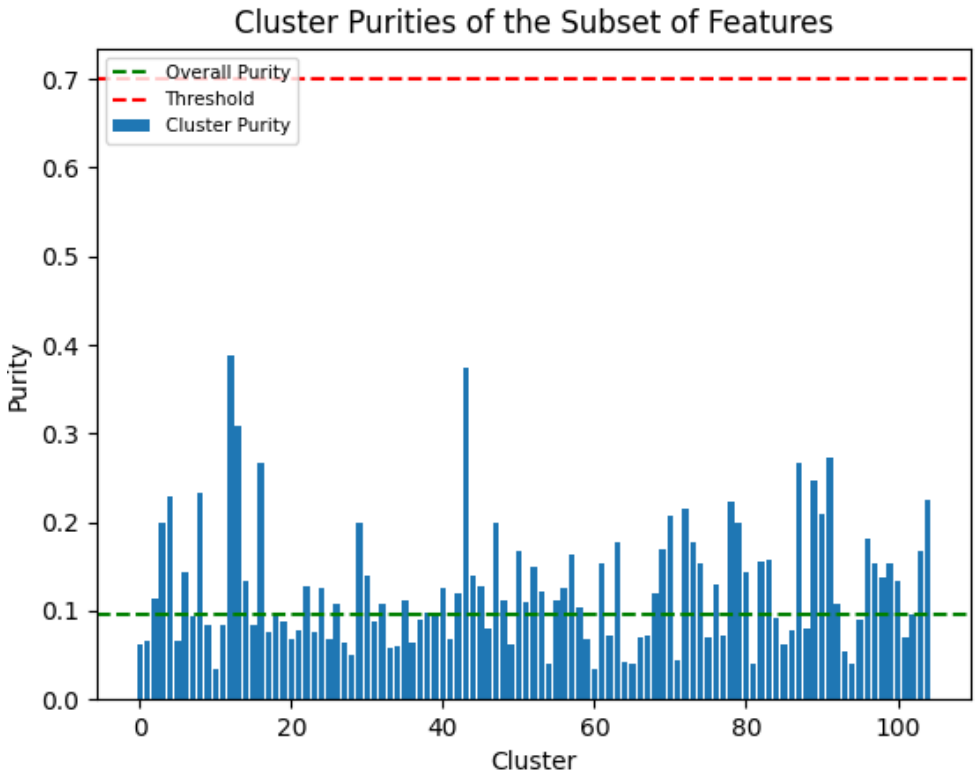


Figure 4.5 – Purity of Each Cluster Using the Optimal Subset of Features (HASC-IPSC dataset).

their privacy is at high risk of being compromised. Assuming that the attacker implements several unsupervised learning attacks, and then integrates and analyzes the results obtained each time, some users can be cross-identified even if there is no information about the users' label. Therefore, if all 219 features are directly transmitted to the cloud, even without the user's label information, their privacy will still be threatened. In contrast, if only the optimal subset of features is transmitted, the purity of the clusters is at most 40%, which poses no threat to the user's privacy.

4.4.4 Experimental Results of WISDM dataset

In addition to the HASC-IPSC dataset, the WISDM dataset is also utilized to validate the robustness of our privacy-preserving feature selection algorithm. To ensure consistency of the initial features, only the smartwatch acceleration data in the dataset are used in our experiments. In addition, due to the similarity of certain activities, they are more difficult to distinguish, so we merge similar activities in the dataset, such as activities

related to eating, thus the final number of activity categories is 12. Furthermore, the maximum number of generations for the NSGA-III algorithm is set to 120 because this dataset is more complex and more generations are needed to obtain a solution that satisfies our conditions. The other parameters remain consistent with the previous settings, and the termination condition follows the second approach. The experimental results of the WISDM dataset are shown in Figure 4.6.

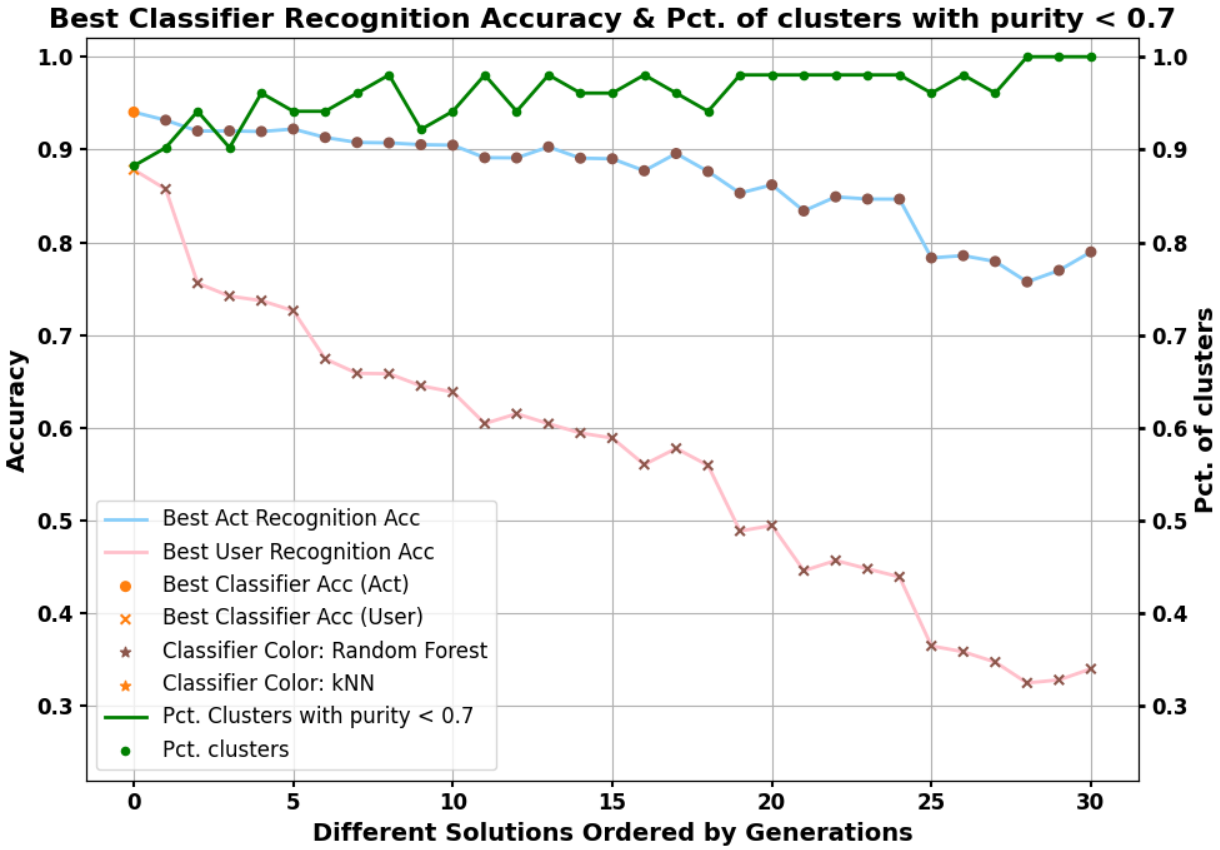


Figure 4.6 – Robust NSGA-III Feature Selection Effects on HAR Accuracy, User Recognition Accuracy, and Cluster Purity of WISDM dataset.

The initial classifier is set to kNN, but as can be seen in Figure 4.6, the best classifier has also been RF since the first generation. Additionally, starting from the 19th solution, the user recognition accuracy reaches our target by dropping below 50%. However, the purity never reaches our goal until the 28th solution (i.e., the percentage of clusters with purity less than 0.7 is 100%). Finally, the optimal solution selected by our algorithm is the 30th. The specific results and the comparison with the recognition results using the

original data and all features are shown in Table 4.3. The feature subset corresponding to the optimal solution is *kurtosis*, *wavelet entropy*, *wavelet energy 2* of the x-axis, *mfcc 11* of the y-axis and *mean difference*, *median absolute difference*, *slope* of the z-axis. If only these features are transmitted, user privacy can be effectively concealed while maintaining relatively high HAR accuracy.

Table 4.3 – Comparison of WISDM dataset recognition results using: Raw Data vs 219 Features vs Subset of Features selected by NSGA-III.

Data Type	**Row (MSENet)**	**219F (kNN)**	**219F (RF)**	**subsetF**
HAR	95.78%	95.29%	94.03%	78.96%
User	88.61%	84.71%	87.82%	34.02%

On the WISDM dataset, we reproduce both semi-supervised attacks (including activity and identity recognition, see Table 4.4) and unsupervised inference attacks (only focusing on identity inference, see Figure 4.7 and Figure 4.8). The results corroborate the findings from the HASC-IPSC dataset, showing that when attackers have access to the complete feature set (i.e., 219 features), they are able to establish connections between users and their data, which could expose other sensitive personal attributes. In contrast, the feature subset selected by our framework successfully blocks such possibilities, limiting the risk of identity inference and subsequent privacy leakage. These results demonstrate not only the effectiveness of the framework in a different data context but also its potential generalizability, as it can provide strong privacy protection across diverse attack strategies and datasets.

Table 4.4 – Semi-Supervised Learning Results of HAR and User Recognition using all 219 features and the optimal subset of features (WISDM dataset).

	Classifier	**DNN**	**RF**	**XGB**
all $219F$	**HAR**	89.31%	88.31%	89.89%
	User	46.43%	63.34%	58.79%
subset F	**HAR**	69.76%	73.37%	73.09%
	User	14.41%	21.03%	18.47%

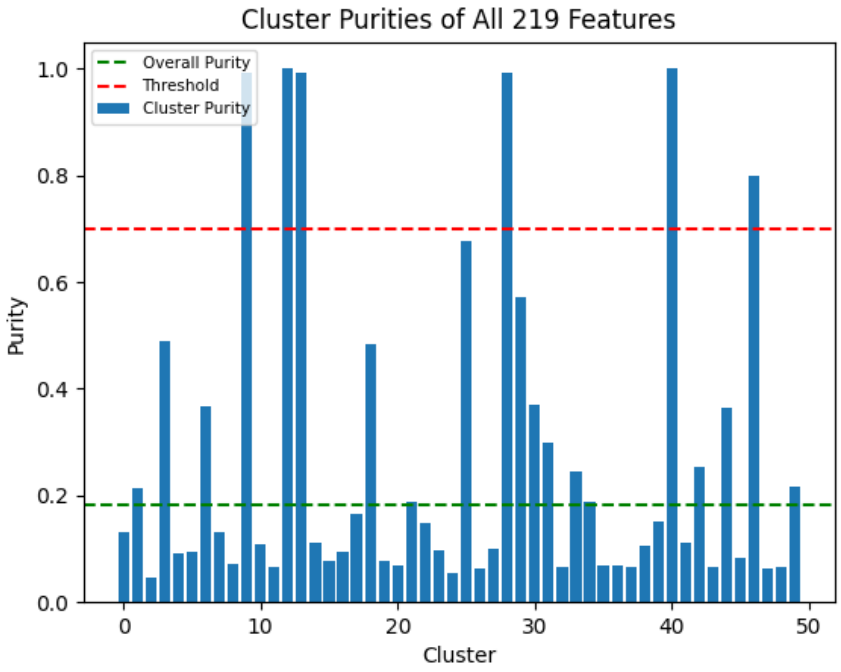


Figure 4.7 – Purity of Each Cluster Using All 219 Features (WISDM dataset).

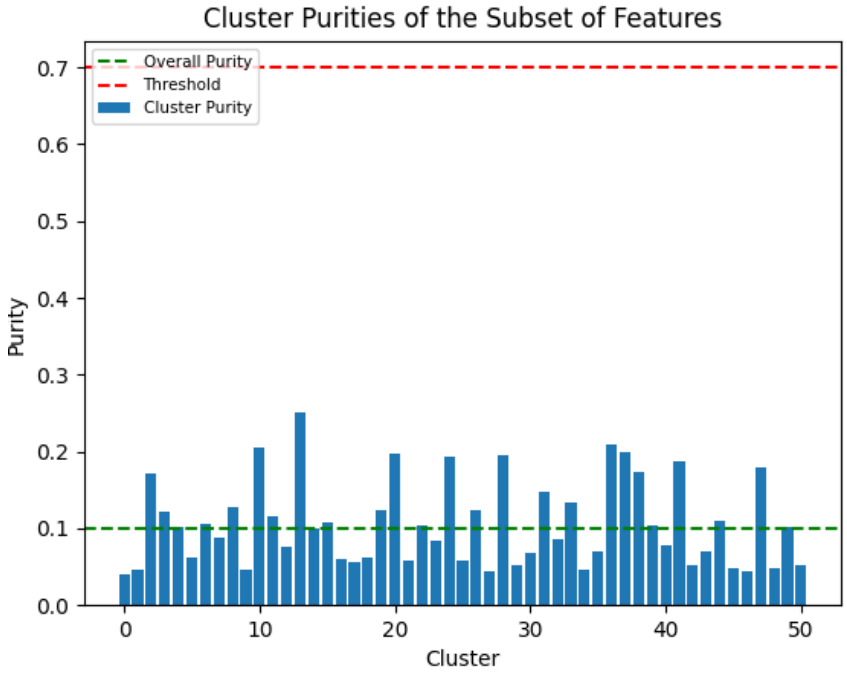


Figure 4.8 – Purity of Each Cluster Using the Optimal Subset of Features (WISDM dataset).

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

4.5 Conclusion

Building on the work in Chapter 3, this chapter proposes a robust privacy-preserving feature selection framework and validates it using two real-world datasets. The robustness of the framework is ensured by comprehensively considering different attacker models and consistently using the best-performing classifier in the feature selection algorithm (i.e., the NSGA-III algorithm). In addition, our approach takes into account the constraints of resource-restricted environments and proposes the corresponding countermeasures. Based on our experimental results, the proposed method can defend against the three attacker models discussed in Chapter 2, which effectively demonstrates the robustness of our framework.

Although the experimental results indicate that the proposed feature selection framework can be deployed on resource-constrained edge servers, demonstrating its applicability and robustness in environments with limited computing and storage capabilities. However, the current experimental validation still has certain shortcomings. In particular, testing of the feasibility of real-time feature extraction and transmission on different types of terminal devices has not yet been conducted. Differences in processing performance, network bandwidth, latency, and power consumption across various devices may impact the algorithm's real-time performance and stability during actual deployment. Therefore, it is imperative to design and implement targeted experiments to comprehensively evaluate its adaptability between devices. It should be noted that, compared to other common privacy protection schemes such as differential privacy and homomorphic encryption, this scheme does not require substantial additional computational or communication overhead during implementation, making it better suited to heterogeneous local environments with a wide variety of terminal devices of varying capabilities. This not only significantly reduces deployment complexity and technical barriers, but also enhances the framework's scalability and promotional value across diverse application scenarios.

Furthermore, Chapter 2 discusses two types of potential semi-supervised learning attacks. In this thesis, the primary focus is on the first type of attack, where the attacker can obtain labels for a small portion of all or most of the user's data. However, from a practical application perspective, the second type of attack may be more common in real-world environments and has a higher probability of occurring. This type of attack involves the attacker obtaining all or most of the data labels for a small portion of users. Therefore, future research will focus on simulating this type of attack scenario and ex-

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

ploring targeted defense strategies to further enhance the robustness and practicality of our feature selection framework in diverse attack environments.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

Chapter 5

InclusiveHAR Dataset

In the previous chapters, we evaluated our approach using two commonly adopted HAR datasets, HASC-IPSC and WISDM, which provide a solid baseline for verifying our ideas. However, these datasets mainly represent conventional HAR scenarios and do not reflect the specific focus of this thesis, namely the privacy of individuals with disabilities. In recent years, a variety of HAR datasets have been released, such as UCI-HAR [221] for general activity recognition and MobiFall [222] for fall detection. Other datasets, including SisFall [223] and HASC-PAC2016 [224], introduced more complex scenarios but were published nearly a decade ago. More recent datasets, such as PAAL ADL [208], consider daily living activities with greater diversity among participants in terms of age and gender. Nevertheless, existing datasets still leave important gaps that limit their representativeness.

Although HAR datasets have grown in number and diversity, almost all publicly available HAR datasets only include data from able-bodied participants. This situation has, to some extent, limited the applicability and broader adoption of HAR technologies across different populations. For example, individuals with motor impairments, limb loss, neurological disorders, or other disabilities often exhibit activity patterns, movement characteristics, and sensor signal distributions that differ significantly from those of able-bodied populations [225]. If models trained on datasets exclusively comprising able-bodied individuals are applied directly to people with disabilities, recognition accuracy and robustness may significantly decline. This not only undermines the practical value of such systems in medical and assistive technologies but may also raise issues of fairness and accessibility. In particular, approximately 1.3 billion people worldwide are affected by some form of disability, representing about 16% of the total global population [3]. However, data specifically designed to reflect the dynamics of this large population is extremely scarce and, in many cases, completely absent. Therefore, in order to better simulate real-life scenarios, it is necessary to consider including a diverse range of participants, especially people with disabilities.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

The purpose of this thesis is to improve the quality of life of people with disabilities and help them achieve greater autonomy and accessibility in a variety of environments while protecting their privacy. Thus, it is urgently necessary to bridge this information gap to promote the implementation of more targeted and effective solutions. Particularly, to the best of our knowledge, there are currently very few datasets that contain data on such groups, e.g., another study conducted by the Kerpape Living Lab team [226] and [225]. This is mainly due to privacy concerns, especially some studies have shown that an attacker can exploit sensor data to recognize sensitive information about the user, such as identity [201], gender, age, height, weight [107], etc. Furthermore, data from these specific groups are more sensitive than data from the general population, because we believe that their data may differ significantly from the data of the general population and, therefore, would be easier to recognize. This hypothesis will be further examined in the data analysis section (Section 5.5). In addition, selecting environments where people with disabilities can participate in experiments, as well as designing the experimental scenarios, require careful consideration.

To address the above shortcoming, this chapter proposes to lay the groundwork for a new HAR-related time-series dataset, which should be understood as an initial step in its construction and has the following notable features. In addition to data from able-bodied participants, it also collects data on the daily activities of a participant with a disability. In fact, the dataset comprises data from four participants (two males and two females, including a person with motor disability). Furthermore, the dataset records a variety of daily activities and is equipped with multi-modal sensor signals and standardized activity labels. During the data collection process, we fully considered the safety, comfort, and privacy protection of the participants to ensure the quality and ethical compliance of the data.

5.1 Related Public Datasets

In this section, we systematically reviewed and analyzed datasets related to HAR published over the past decade. This review not only helps to gain a deeper understanding of the current state-of-the-art and evolving trends in this field in terms of data collection techniques, sensor configurations, annotation strategies, open policies, and application scenarios, but also reveals the differences and commonalities among different datasets in terms of scale, diversity of participants, coverage of activity categories, and data quality

control. A comprehensive understanding of these characteristics not only provides valuable insights for the design philosophy, data collection strategies, and experimental design in this chapter, but also lays a forward-looking foundation for subsequent data processing workflows. This enables the identification and mitigation of potential issues in data cleaning, format conversion, quality assessment, privacy protection, and bias correction, thereby ensuring that the released dataset achieves a high standard in terms of scientific rigor, usability, and reproducibility, and possesses the potential to drive further research and application in the field of smart healthcare monitoring.

It should be noted that our search for public datasets focused on sensor datasets, i.e., time series data measured using wearable devices. In addition, some basic requirements were set when filtering the datasets: the raw data must be freely available and the datasets must contain Activities of Daily Living (ADLs). Based on this premise, the selected datasets are broadly categorized into two groups: fall detection related and human activity recognition related.

— MobiFall [222]: 24 volunteers between 22 and 47 years old. They performed nine daily activities and four fall tests using a Samsung Galaxy smartphone. Among them, 15 subjects performed only the fall test. Data from three sensors were included: accelerometer, gyroscope, and orientation. In addition, besides activity labels, the dataset also contains subject id, age, gender, height and weight information of the volunteers.

— SisFall [223]: 38 participants recorded 19 types of ADLs and 15 types of falls with a self-developed embedded device fixed to their waist. They were divided into two groups: young adults (19 to 30 years old), and elderly people (60 to 75 years old). The elderly people did not perform falls and four types of activities. Data were acquired with three sensors (2 accelerometers and 1 gyroscope). This dataset also contains participant labels and their information (age, height, weight, gender).

— UCI-HAR [221]: 30 volunteers aged between 19 to 48 years performed six types of ADLs. Data were collected using a Samsung Galaxy S II smartphone equipped with accelerometer and gyroscope sensors. In addition to activity labels, the dataset also includes subject id information.

— HASC-PAC2016 [224]: 510 participants (from early 20s to late 60s) recorded six basic ADLs. Data were captured using smartphones (Samsung and Sony) equipped with multiple sensors: accelerometer, gyroscope, magnetometer, location sensor, barometric pressure sensor, proximity sensor, light sensor, and WiFi sensor. In this dataset,

there are also subject labels and information about gender, age, height and weight.

— PAAL ADL [208]: 52 healthy subjects (18 to 77 years old) performed 24 different ADLs. During the experiment, participants wore an Empatica E4 bracelet with an accelerometer on the wrist of their dominant hand. In addition to activity labels, the dataset contains participant id, gender and age information.

— WISDM [196]: 51 subjects performed 18 diverse ADLs. Data were collected from the accelerometer and gyroscope of a smartphone (Nexus 5, Nexus 5X, or Galaxy S6) and smartwatch (LG G watch). This dataset includes subject id and activity labels.

Compared with the dataset constructed in this chapter, existing HAR datasets generally lack coverage of special groups, such as people with physical disabilities, resulting in insufficient representation of population diversity. In addition, some datasets are relatively limited in their design of activity categories, failing to cover a wider range of representative daily activities or complex situations, thereby limiting their application and model generalization capabilities.

5.2 Experimentation Setup

The Centre Mutualiste de Rééducation et Réadaptation Fonctionnelles de Kerpape (Centre de Kerpape) is an institution that integrates rehabilitation, research, and social reintegration, relying on a multidisciplinary team to provide individuals with disabilities support for physical and mental recovery, as well as social reintegration. The Centre de Kerpape is a key partner in the Handicap Innovation Territoire (HIT) project (2020–2027, initiated by Lorient Agglomération, Centre de Kerpape, and Biotech Santé Bretagne), which aims to promote regional inclusiveness and technological innovation through issues related to disability. Within the HIT framework, the Centre de Kerpape participates in identifying the needs of people with disabilities and testing innovative solutions for technology, services, and organizational models in real-life environments. We collaborate with the Centre de Kerpape and its staff to collect data, with the data collection environment being an apartment provided by the Centre de Kerpape for people with disabilities to reside in.

Specifically, our experiment was carried out in a specially designed smart apartment equipped with various sensors and controllable devices, aimed at providing a convenient

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

living environment for people with disabilities and supporting their independence in daily life. The floor plan of the apartment is shown in Figure 5.1. The apartment comprises a bedroom, a bathroom, and an open-plan space that combines kitchen and living room functions, simulating a real-life home environment. The bed in the bedroom can be adjusted in height and angle via a tablet or remote control to accommodate users with different physical conditions; blinds, doors, and lighting systems can also be remotely controlled, ensuring residents can easily perform on/off operations. The kitchen and living area are equipped with smart lighting, etc., making cooking, cleaning, and other daily activities more convenient.



Figure 5.1 – Floor plan of the apartment used in the experiment.

The apartment is equipped with various sensors, including motion sensors and environmental monitoring devices, to record residents' activities in real time, such as lighting and blind usage. Sensor data related to residents' activities in the experiment, such as walking and standing up, was recorded via a tablet. All devices are connected to a central control system and can be managed uniformly via a tablet, enabling monitoring and recording of residents' operations. The layout and facility design of the apartment not only simulate everyday living scenarios, but also provide detailed and objective data collection conditions for research, laying the foundation for assessing the feasibility of different assistive devices and living-environment designs.

In our experiment, the tablet (as shown in Figure 5.2) is primarily used to control the smart apartment, specifically through an augmented reality application. For this reason, it also serves as a capture device, enabling data collection without relying on a sensor fixed to a harness or belt. Through the Artefacto application (see Figure 5.2) running on the tablet, volunteers can manually control indoor smart devices, as shown in Figure 5.3, including lighting, doors, and other environmental control devices. During the experiment, volunteers naturally held the tablet in their hands according to their own usage habits to ensure the naturalness of the interaction process and the convenience of operation, thereby obtaining more realistic behavioral data.

The experiment was conducted primarily in three areas of the apartment: the bedroom, living room, and kitchen, to fully cover the main scenarios of activity in daily life. The entire experiment was recorded on video to ensure complete capture of the participants' behaviors and device operations, providing reliable video material for subsequent data processing, behavioral analysis, and result verification. In addition, the video data can also be used to cross-verify the data recorded by the sensors, improving the integrity of the experimental data and the accuracy of the analysis.

5.3 Activities Setup

In this chapter, two main experimental scenarios were designed to systematically collect various human motion data: the first is "door passing" and the second is "activities of daily living (ADL)". Each scenario includes two different experiments to ensure the richness and representativeness of the data. For the door-passing scenario, the specific experiments were designed as follows: 1) *open door 1, open door 2, go out*; 2) *open door 1, open door 2, close door 1, go out*. In the ADL scenario, we focus on individual be-

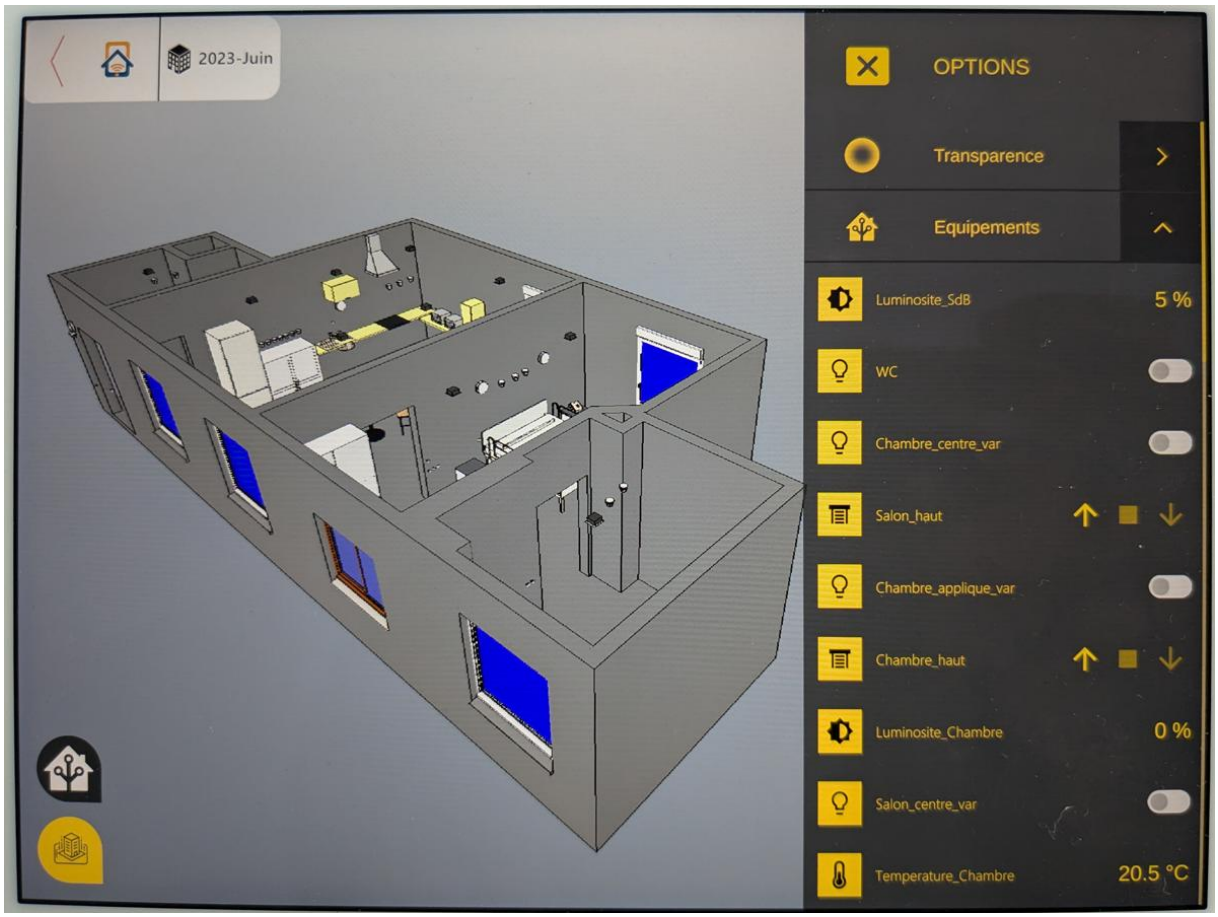


Figure 5.2 – Tablet Equipped with the Artefacto Application. The tablet included an accelerometer, a gyroscope and a magnetometer.

havior when performing daily tasks in a natural state. The types of ADL tasks and their descriptions are listed in Table 5.1. To ensure the reproducibility of the study and the standardization of data management, we have created separate folders for each scenario under the data directory and provided an *scenario.txt* file within each folder, which details the scenario configuration and the equipment specification for subsequent analysis and verification.

In order to simulate real life more realistically, participants performed activities naturally in the environment and did not receive instructions on how or for how long to complete each activity. As shown in Figure 5.3, due to the augmented reality application, the device was held in the hands of the participants during the experiment.



Figure 5.3 – Experimental Scenario Illustration.

Table 5.1 – The activities represented in the dataset and their descriptions.

Activity	**Description**
lights_on_br	Turn on the bedroom light
blinds_open_br	Open the bedroom blinds
walk	
lights_on_kit	Turn on the kitchen light
bowl_pick	Take a bowl from the table
microwave_bowl_in	Put the bowl into the microwave
tv_on	Turn on the TV and select the desired program
blinds_close_lr	Close the living room blinds
microwave_bowl_out	Take the bowl out of the microwave and place it on the table

5.4 Time-series Sensor Data

The raw time-series sensor data were recorded by the tablet. Sensor data is recorded each time the sensor is updated (updated value) and according to the application fluidity. As these data is very noisy, updates are frequent, so the recording frequency is linked to the application's update loop. This update loop integrates calculations and 3D rendering, and is thus dependent on the data displayed and the capacity of the hardware running the application. Depending on the point of view, the number of frames per second may vary

between 15 and 30 Hz. Therefore, there is no fixed frequency sampling of these sensor data, a limitation that will be further discussed later in this chapter.

Our experiments recorded the following data: acceleration, gravity, and rotation associated with HAR. These data were stored in the folders *passing_through_doors* and *daily_activities* according to the two scenarios, respectively. Within these two folders, data from different participants were stored in their respective folders, e.g., all data from the first participant were stored in the *person1* folder. There are seven files in each person's folder, i.e., six data files and one personal information file for each participant. Each scenario was designed with two experiments, and each experiment had three HAR-related files. The sample filenames are as follows.

— `Exp1_Accel.csv`

— `Exp2_Gravity.csv`

The first component identifies the experiment id, which is either "Exp1" or "Exp2". The second component is "Accel" or "Gravity" or "Rot" to identify the type of data. All data files end with ".csv", and personal information files end with ".txt".

In time series data files, each raw contains one sensor reading. The format of this line varies depending on the experimental scenario. In general, each raw of the HAR-related data file is formatted as a timestamp and xyz-axis data. However, the door-passing scenario lacks information on the activity label. Therefore, for the daily activity scenario, the format of each line of the HAR related files is: *Timestamp, X, Y, Z, Activity*. While for the door-passing scenario, the format of each line of the HAR related files is: *Timestamp, X, Y, Z*. An example of an acceleration data file for the ADL scenario is shown in Table 5.2.

Table 5.2 – Example of an acceleration data file for the activities of daily living scenario.

Timestamp	**X**	**Y**	**Z**	**Activity**
1718113649017	-0.008113647	-0.001839246	0.006952816	walk
1718113649051	-0.004303834	0.001388989	0.005331323	walk
1718113649080	-0.00659501	-0.006902643	-0.006020001	walk
1718113649119	-0.00772424	-0.006811813	-0.009423967	walk
1718113649151	-0.003262053	-0.007792751	-0.009365501	walk

In this experiment, in addition to activity data closely related to HAR, we also recorded

trajectory data related to the generation of indoor floor plans. However, due to the relatively limited number of participants in this experiment, the trajectory data obtained may have certain limitations in terms of scale and diversity, and were not exploited in the present analysis. Nevertheless, in consideration of the integrity of the research design and methodological foresight, a brief introduction to this type of data is still provided here to lay the foundation for future research. In the future, we plan to continue refining and systematically conducting related experiments under the premise of further expanding the scale of participants and optimizing the experimental process, thereby obtaining more representative and universally applicable data to provide reliable support for in-depth exploration of trajectory data in spatial modeling and application research. In addition, the trajectory data was recorded every 100 ms (i.e., 10 Hz). The format of each row of the floor-plan generation-related file is: *Timestamp, PosX, PosY, PosZ, Pitch, Yaw, Roll*. Table 5.3 presents an example of this type of data file to illustrate the organizational form of the experimental data.

Table 5.3 – Example of a trajectory data file.

Timestamp	**PosX**	**PosY**	**PosZ**	**Pitch**	**Yaw**	**Roll**
1694431187335	17.74383	1.297997	-17.21651	62.41424	285.5115	359.2865
1694431187443	17.74643	1.293666	-17.21505	62.6024	284.7437	358.4351
1694431187575	17.74683	1.289903	-17.21641	62.63164	284.1016	357.8454
1694431187708	17.74366	1.289622	-17.22287	62.32595	283.7951	358.0529
1694431187841	17.74077	1.291883	-17.2319	61.30852	281.8661	357.0779

5.5 Analysis of Time-Series Sensor Data

In order to more clearly compare the time series data of people with disabilities and individuals without disabilities as an illustration of the importance of data privacy protection, the acceleration data from the second experiment of the door-passing scenario are selected for analysis in this section. In addition, we randomly select data from an able-bodied participant in addition to data from the individual with a disability in the dataset. Finally, a brief introduction will be given on how to analyze trajectory data for spatial modeling (i.e., the generation of indoor floor plans).

5.5.1 Raw Data Comparison

Raw acceleration data for the person with a disability and an able-bodied individual in the door-passing scenario are shown in Figure 5.4. As can be seen from the figure, the acceleration data in either axis fluctuate more, and the noise is more pronounced for the participant with a disability. In contrast, the data of the able-bodied participant are more stable and periodic. This phenomenon may result from an increase in gait instability in the participant with a disability. In addition, for these two participants, the data for both x-axes are a bit more stable compared to the other two axes. However, such a comparison has already revealed the differences between their data, and therefore it is necessary to implement privacy protection measures before releasing the data in the HAR system.

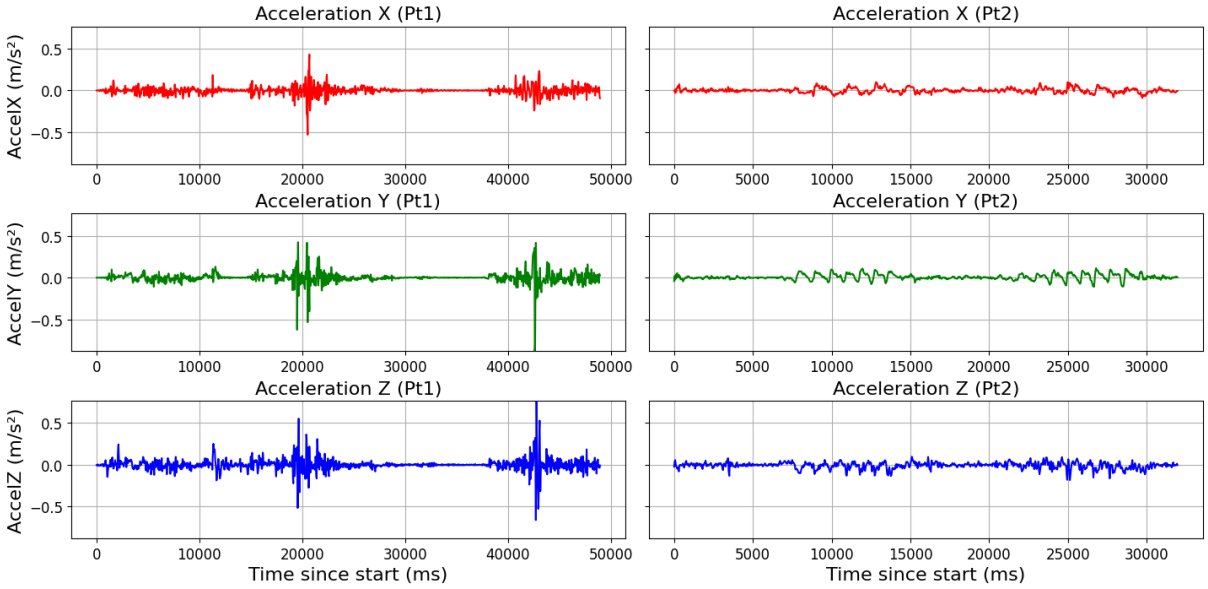


Figure 5.4 – Triaxial acceleration recordings during door-passing scenario by the participant with a disability and an able-bodied participant.

5.5.2 Statistical Information Analysis

This subsection will further analyze the data of the two participants, while the previous subsection showed their raw data, this subsection analyzes the features of their data, for this purpose, the following features are extracted from their raw data: mean, standard deviation (std), root mean square (rms), range (max - min), jerk std, spectral entropy,

and energy. The size of the sliding window is 45. In addition, calculations confirm that both datasets were sampled at 30 Hz. The results are shown in Figure 5.5.

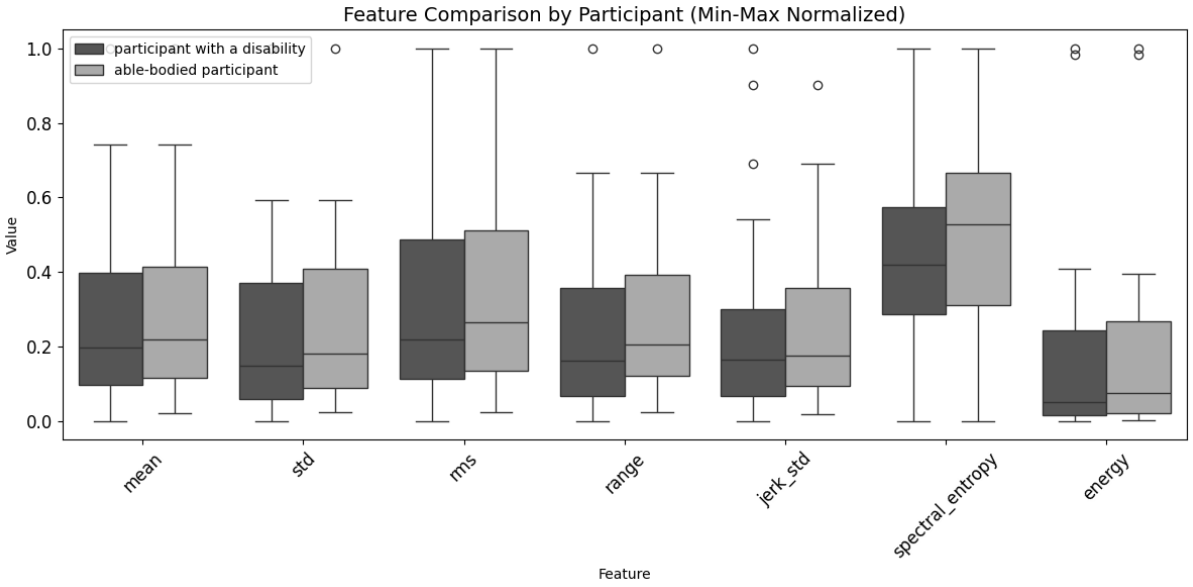


Figure 5.5 – Comparison of Statistical Information on the participant with a disability and the able-bodied participant.

The mean, std, rms and range are all lower for the participant with a disability compared to the able-bodied participant. This suggests that the individual with a disability exhibits a lower overall activity intensity and amplitude, as well as fewer variability of acceleration (movements are slower). Furthermore, the able-bodied participant's jerk std, spectral entropy, and energy are higher. A higher jerk std indicates that the able-bodied individual exhibits more intense motion variations, leading to greater activity intensity and thus higher energy. He/She presents a higher and more widely distributed spectral entropy during movement, which may indicate greater flexibility and less regular motion patterns.

However, it should be emphasized that the current statistics are by no means representative. They are provided solely to help formulate preliminary hypotheses, as no definitive conclusions can be drawn from such a limited amount of data.

5.5.3 Trajectory Data Analysis

Trajectory data, as a typical form of spatiotemporal data, is widely used in research fields such as human behavior modeling, indoor navigation, path optimization, and environmental understanding. In this chapter, the collected trajectory data includes timestamps, spatial position information (PosX, PosY, PosZ), and posture parameters (Pitch, Yaw, Roll). This data not only accurately reflects the movement trajectories and directional orientation of participants during the experiment, but also provide a foundational basis for subsequent environmental modeling and spatial semantic extraction. In other words, this data primarily reflects the participants' movement paths and spatial distribution characteristics during the experiment, which aims to provide a preliminary reference for subsequent environmental modeling and spatial understanding research based on behavioral trajectories.

Combined with Figure 5.6, the walking paths and areas of stay of a participant in the experimental apartment can be clearly observed. By analyzing the spatial distribution of the trajectories, features such as room boundaries, structural connections between spatial areas, and frequency of space use could be identified. For example, the aggregation of continuous trajectory lines at the edges of a space often indicates the presence of obstacles (e.g., the walls), while the lingering and dense distribution of trajectories in a particular area may correspond to functional zones (e.g., the kitchen). Furthermore, by combining posture information (Pitch, Yaw, Roll), it may be inferred that the participants' perspective and direction of focus within the scene, thereby incorporating additional semantic interpretations into the spatial modeling process.

However, the value of trajectory data also comes with potential privacy risks. Since trajectories record an individual's complete movement path in a specific environment, they can not only reveal users' spatial activity patterns but also, to some extent, infer their behavioral habits, lifestyle patterns, and even personal information (e.g., the disability status). When combined with spatial layout data, the inferred information may be more sensitive than the original trajectory data itself. For example, by analyzing the points where users linger in their trajectories, their primary activity areas can be exposed; by identifying repetitive action patterns, their daily routines and interests can be inferred. Therefore, while using trajectory data for indoor modeling and spatial understanding, it is essential to fully address potential privacy risks and implement the corresponding anonymization and protection measures during data collection, storage, and use. This could constitute future research and is not covered in the present thesis.

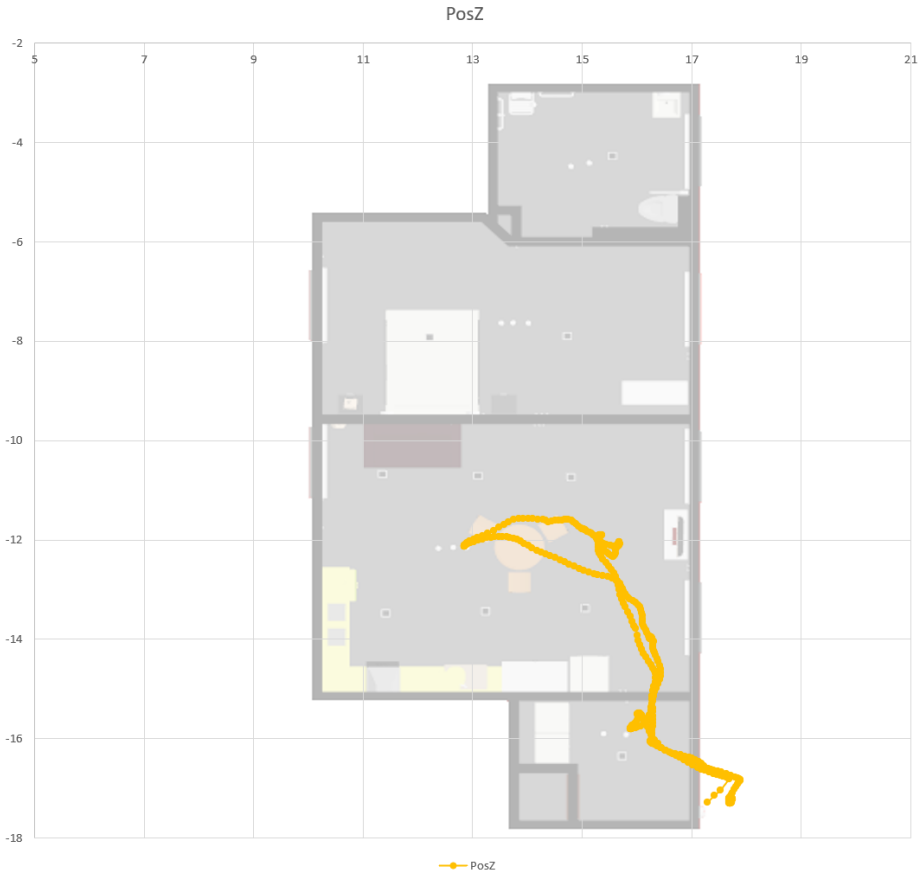


Figure 5.6 – Example Floor Plan with Participant Trajectories derived from collected data.

5.6 Private Information Recognition Experiment

In this section, a preliminary analysis of our dataset will be provided, although this dataset does not contain much data, it is interesting to analyze it as a novel dataset in the field of HAR. We process the data according to the common pipeline: preprocessing, feature extraction, and classification. This chapter considers two attacker scenarios discussed in Chapter 2: 1) attackers with background knowledge could access all data and know the labels of some of it; 2) ordinary attackers may be able to only obtain all data. Therefore, two different types of classification model are employed in our recognition experiments, i.e., supervised learning models and unsupervised learning models.

5.6.1 Experiment with Supervised Learning Models

First, we focus on the recognition results of the supervised learning model. Since the sampling frequency of the sensor is not fixed (it varies from 15Hz to 30Hz), in order to achieve balance, the acceleration, gravity, and rotation data are segmented into 45-reading windows with a 0.5 overlap. The overlap step aims to minimize information loss in the data segmentation process. Subsequently, all data are normalized to prevent the effects of different data ranges on the results.

In the feature extraction step, according to Chapter 4, 219 features are extracted from each type of data separately. Therefore, finally, 219 * 3 = 657 features are extracted from the data. In this chapter, the following machine learning models are used for the classification task: Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), XGBoost (XGB) and Random Forest (RF). Their respective parameters are as follows.

— SVM: it uses the RBF kernel, with C = 50, gamma = *'auto'*, and class_weight = *'balanced'*

— k-NN: the parameters are neighbors = 1, leaf-size = 1, and p = 1

— XGB: the parameters used are n_estimators = 400, learning_rate = 0.1, max_depth = 6, min_child_weight = 3, gamma = 0.2, subsample = 0.8, colsample_bytree = 0.8 and reg_lambda = 1.0

— RF: it uses n_estimators = 100, min_samples_leaf = 1, max_depth = 50 and min_samples_split = 2

Unlike previous common HAR datasets, we pay more attention to the identification of users' private information in addition to activity recognition. Especially for people with disabilities, since their information is more sensitive and their data seem to be different from the data of able-bodied individuals, their data are more threatened and more likely to be recognized. In our recognition experiments, the scenario of passing through the door does not involve different activities, so only the recognition results of sensitive information are available, whereas the recognition results of daily life data include activity recognition and user sensitive information recognition. These results are shown in Tables 5.4 and 5.5, respectively.

Based on the results in the tables, it can be seen that the overall accuracy of recognizing users' personal information is much higher than that of activity recognition. This may be due to the fact that there are only four participants, and therefore activity recognition is more complex compared to the other classification tasks. However, even though we do not

Table 5.4 – Recognition results of door-passing scenario.

Classifier	**SVM**	**k-NN**	**XGB**	**RF**
Participant	97.96%	96.94%	100%	97.96%
Disability	97.96%	100%	100%	96.94%
Gender	100%	100%	100%	100%

Table 5.5 – Recognition results of daily life scenario.

Classifier	**SVM**	**k-NN**	**XGB**	**RF**
Participant	96.88%	97.32%	100%	97.77%
Disability	99.11%	100%	100%	99.55%
Gender	98.21%	99.11%	99.55%	97.77%
Activity	70.09%	82.59%	70.09%	77.23%

have much data, different activities can still be recognized with relatively high accuracy. Furthermore, the recognition accuracy is very high for the user's private information, whether it is the user identity, the presence of a disability, or the gender. Especially for relatively simple binary classification tasks, such as disability status recognition and gender recognition. The reported near-100% accuracy for recognizing identity, gender, and disability status suggests that the captured motion signatures are sufficiently distinctive to enable fine-grained identification of these personal attributes. For recognition of disability status, the wheelchair participant might exhibit more abrupt or constrained gestures, or maintain a more stable grip on the tablet due to reduced limb mobility, making him/her easy to distinguish from other participants. In addition, in our experiments, the best performing classifier in the activity recognition task is k-NN, whereas the best performing classifier in the user private information recognition task is XGB. This shows that user privacy in the HAR system is under significant threat, especially for special groups, e.g., people with disabilities. Therefore, before releasing the HAR data, the necessary privacy measures need to be implemented to protect user privacy from various potential attackers.

5.6.2 Experiment with Unsupervised Learning Models

In this subsection, the results of the unsupervised learning model are discussed. Data preprocessing and feature extraction are consistent with that in the previous subsection. However, here k-means is used for data clustering, and subsequently the confusion matrix is calculated to measure the clustering performance of k-means. In addition, it is important to note that mixing data from different activities affects the effectiveness of the clustering, so in this part of the experiment only data from the door-passing scenario are used. Furthermore, since among all personal information of the user, we are more concerned about whether the disability attribute can be identified, this subsection only analyzes the grouping results of the disability attribute, as shown in Figure 5.7. The *n_clusers* of the k-means algorithm is set to 2 because there are only two categories.

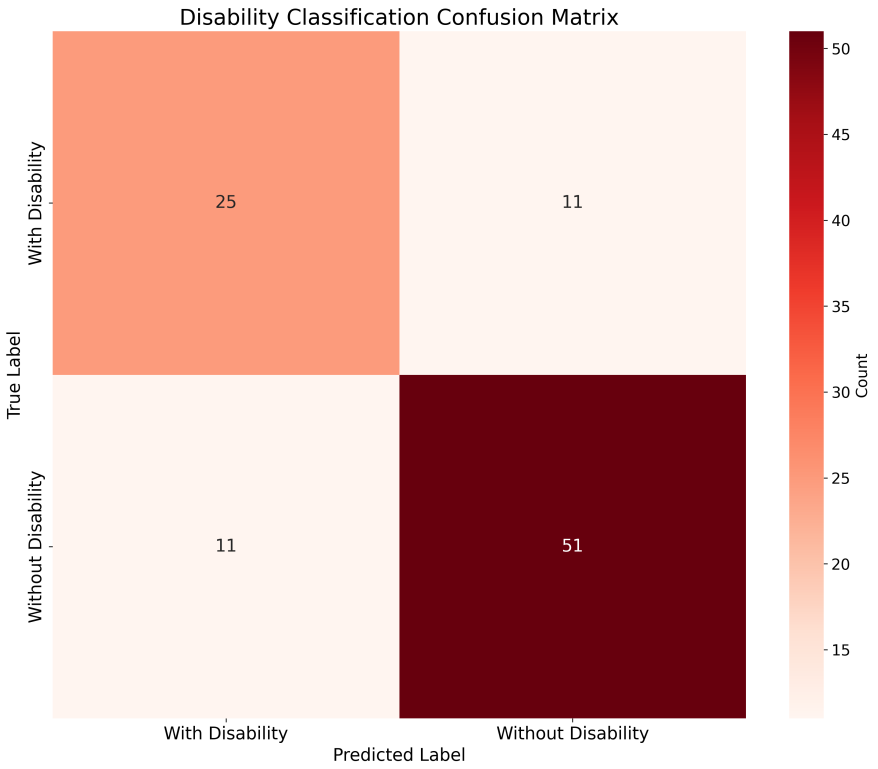


Figure 5.7 – Confusion Matrix for Disability Classification Using Data from the Door-Passing scenario.

Based on the confusion matrix, if the purity of each cluster is calculated, i.e., the number of correct predictions/total number of predictions, a purity of 69.44% for the disability attribute and 82.26% for the non-disability attribute can be obtained. In other

words, even without labels (with or without disability), a malicious attacker may be able to cluster data with the help of unsupervised learning models. Moreover, since this is a binary classification problem, the attacker is likely to recognize most of the data with disability attribute (or without disability attribute), which leads to a threat to the privacy of people with disabilities.

5.7 Conclusion

As wearable devices become increasingly prevalent in our daily lives, such as fall monitoring for the elderly. Researchers are interested in recognizing activities using data measured by these devices. As a result, many HAR-related datasets have been published in the last decade. However, as far as we know, very few datasets contain data from people with disabilities. For this reason, we laid the foundation for the InclusiveHAR dataset in this chapter, which comprises sensor data from both people with disabilities and able-bodied individuals. Our experimental results indicate that there is a high probability that users' private information will be identified. Furthermore, even without label information, unsupervised learning methods can still be used to distinguish disability attribute data from non-disability attribute data, which poses a great threat to the privacy of people with disabilities.

Therefore, this work should be regarded as a preliminary step. The top priority for future research is to expand and diversify the dataset in order to validate our hypotheses. This involves designing more complete scenarios, recruiting a broader range of participants, and in particular enriching the disability category. Such efforts will not only enlarge the dataset but also enhance the trajectory data, thereby providing a stronger basis for analysis. The framework discussed in Chapter 4 remains a possible solution for protecting user privacy in real-world HAR systems, but its validation requires richer and more representative data. Finally, the work in this chapter highlights the importance of taking diversity into account in the construction of HAR datasets, both by including people with disabilities and by considering other potentially vulnerable minorities.

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

Conclusion

The rapid growth of IoT technologies, whether in a private or industrial context, allows for the acquisition of a multitude of data that can be used to feed increasingly large knowledge bases (Big Data). Among all this data, some is already considered sensitive because it is directly linked to personal information and generally requires the user's explicit consent. This is the case, for example, with GPS coordinates or access to the camera or microphone. However, other sensors (such as accelerometers or gyroscopes) are also present, but their impact on privacy or security is less obvious. Indeed, certain private or secret information about the users could be inferred from them, such as weight, height, age, and gender [107], and even identification of different users [200, 201, 202]. Furthermore, in this case, it is potentially the fact that they are cross-referenced with other information that poses a risk to privacy.

Within this context, this thesis is dedicated to ensuring the privacy and security of sensitive health-related data from people with disabilities. In fact, measures taken to facilitate the lives of persons with disabilities (e.g., healthcare monitoring systems, smart devices, etc.) may result in the extraction of information related to the identity, privacy, or disability of these populations (as shown in Chapter 5). The primary objective of this thesis is to identify and classify information that can be inferred from data captured (i.e., time series data) by IoT devices, traditional mobile devices, and smart home devices specific to HIT projects, which is related to privacy of individuals with disabilities.

In the first chapter, we conducted a systematic and comprehensive review of the potential privacy threats and attack methods faced by healthcare data, with a particular focus on the assisted living scenarios. This analysis went beyond a simple enumeration of risks: it carefully examined how different types of adversaries could realistically exploit data at various stages of processing, and what kinds of sensitive information could be inferred as a result. At the same time, we explored existing protection strategies and critically assessed their applicability to assisted living contexts. This work not only lays a solid foundation for understanding the broader research context and academic landscape but also provides essential theoretical support for the subsequent proposal of privacy-preserving solutions introduced in this thesis.

Based on this foundation, we constructed an attacker model tailored to real-world application scenarios. This model specifies the potential roles, objectives, and capabilities of adversaries that may realistically arise in Human Activity Recognition (HAR) systems. Such a formulation enables a more precise characterization of potential threats and provides guidance for designing privacy protection mechanisms. Building upon this model, we then proposed a privacy-preserving feature selection scheme designed for deployment in local environments, such as edge servers working in conjunction with user devices. The central aim of this scheme is to mitigate privacy leakage risks as early as possible, before sensitive information leaves the local domain, while at the same time ensuring that the selected features retain sufficient discriminative power to support reliable activity recognition. This dual objective is crucial: protecting privacy should not come at the expense of rendering data ineffective, and our approach explicitly balances these two requirements. Furthermore, by progressively expanding the assumptions regarding attacker types and capabilities, we refined and optimized the robustness of the proposed framework. Instead of limiting the analysis to a single, simplified adversary model, we considered a variety of more complex and diverse scenarios, which allowed us to test the resilience of our robust approach under more challenging conditions.

Finally, considering that current research in this field largely relies on common public datasets, it is important to note that while these datasets are widely used for benchmarking and facilitate comparability across studies, they often lack adequate coverage of special groups, particularly people with disabilities, whose data may present very different patterns. This shortcoming has important consequences, as it may restrict the universality, fairness, and applicability of existing privacy-preserving methods when deployed in real-world environments where user populations are more heterogeneous. To address this issue, we collaborated with the Centre de Kerpape to design and collect a novel HAR-related dataset that includes participants with disabilities. The construction of this dataset serves a dual purpose: it not only provides a more representative and challenging data foundation for future privacy protection research but also contributes to the inclusiveness of HAR studies by ensuring that vulnerable populations are not overlooked. However, it should be noted that this represents an initial step, as the amount of data collected remains limited due to the relatively small number of participants.

In summary, this thesis has made preliminary contributions in the areas of local privacy protection solutions and attacker model design, and has proposed some insightful ideas. However, we recognize that these contributions have also raised new questions and opened

avenues for further improvement. Based on this, the subsequent sections of this chapter will further discuss potential future research directions and work prospects, with the aim of providing a reference for the continued development of the relevant field.

Challenges in Implementing Local Privacy-Preserving Solutions

As the demand for privacy protection continues to grow, researchers have increasingly explored the deployment of privacy-preserving solutions in local environments and even their application in real-time systems. In this context, our feature selection scheme represents a suitable and relevant approach; however, we did not conduct a direct comparison with alternative approaches such as those based on Differential Privacy (DP), Federated Learning (FL) [108], or other machine learning or deep learning methods [106]. Addressing this gap could be an interesting direction for future work.

In addition, the limitations in computational power, storage space, and energy consumption of local environments and devices remain significant obstacles to implementing privacy-preserving solutions. Our proposed approach is no exception: while our experiments suggest that it can in principle be deployed on an edge server, it has not yet been validated in real-world scenarios. Moreover, the feasibility of performing real-time feature extraction directly on diverse IoT devices, which is critical for practical deployment, also remains untested. These gaps highlight important challenges and indicate promising directions for future work.

Furthermore, the diversity of data types poses additional challenges for local privacy protection. In this thesis, we have focused primarily on time-series data, which is the core modality in most HAR systems. However, in practical assisted living scenarios, other modalities such as speech or video data may also be combined with time-series signals to better monitor and support people with disabilities. In such cases, the challenge extends beyond single-modality protection to ensuring the adaptability and universality of privacy-preserving methods across different data types, an issue that remains largely open for future exploration. Existing studies [227, 228] have proposed several solutions to this problem, but further work is still needed to improve and extend them.

Moreover, in HAR systems, privacy protection often relies on predefining which types of information should be hidden. In this thesis, for example, we focused on concealing identity-related features. However, different users may have different expectations: some

Ensuring privacy protection in data outsourcing by mobile applications an IoT Qiyang Li 2025

may prefer to hide gender information, others may prioritize concealing health conditions, and people with disabilities may be most concerned about protecting information related to their disability. This variability in user preferences introduces additional challenges for the real-world deployment of privacy-preserving solutions, as a single predefined strategy may not align with the diverse needs of all users.

To conclude, although existing research has made initial progress in local privacy protection, its practical application remains challenging due to limitations in device capabilities, the presence of multiple data modalities in HAR systems, and the diverse privacy preferences of users. In this thesis, we proposed a feature selection scheme as a possible solution, but did not conduct comparisons with other approaches such as DP-based methods. In order to advance the transition of local privacy-preserving solutions from theory to practice, future research needs to explore algorithm optimization, cross-modal adaptability, personalized privacy requirements, and more extensive experimental validation.

HAR Data: Exploring the Depth and Breadth of Dataset Diversity

In existing HAR research, the diversity of datasets is both a key driver of model generalization and application expansion, as well as a challenge that needs to be addressed. On the one hand, most of the current public datasets focus on a limited number of activity types, standardized collection environments, and relatively homogeneous participant groups, which limits the applicability of models in complex real-world scenarios.

Future research should focus more on "depth" diversity, i.e., exploring finer-grained differences within existing activity categories. For example, the same "walking" action may exhibit significant differences in movement patterns, rhythms, and sensor signal distributions between young and elderly individuals, as well as between healthy individuals and rehabilitation patients. Depth diversity should also encompass subtle variations at the contextual level, such as how the same action is performed under different ground surfaces, weight-bearing conditions, or emotional stress. This "internal diversity" helps train more sensitive and adaptive models, thereby avoiding recognition bias caused by over-reliance on a single pattern.

At the same time, diversity in terms of "breadth" is equally indispensable. Breadth not only means expanding the types of activities to cover more everyday and non-everyday behaviors (such as sports, occupational movements, and responses to emergencies), but

also includes expanding the population and scenarios. In particular, coverage of special groups (e.g., children, the elderly, people with disabilities, and patients with chronic diseases) can compensate for the current research's overemphasis on healthy adult samples. Furthermore, breadth should also be reflected in cross-cultural and cross-geographical considerations, as differences in lifestyle habits, environmental facilities, and even wearing habits may lead to systematic shifts in behavioral signals.

Therefore, future data construction still needs to strike a balance between scalability and fairness. How to establish large-scale open datasets that cover multiple dimensions, multiple populations, and multiple scenarios while protecting privacy and reducing collection costs will directly determine the depth and breadth of HAR technology implementation in applications such as smart healthcare, rehabilitation assistance, and smart homes. Exploring systematic methods for data diversity is not only an important direction for enhancing algorithm robustness and fairness, but will also become a core challenge in advancing HAR research toward widespread applicability in the real world. We therefore hope that the inclusiveHAR dataset developed in this thesis will provide a foundation for future work in this area.

Comprehensive Services for People with Disabilities: Balancing Care and Privacy Protection

In the process of building comprehensive services for people with disabilities, balancing high-quality care with rigorous privacy protection remains a long-term and complex task. The previous chapter discussed privacy challenges in different scenarios (i.e., the three attacker models) and proposed our solutions. However, it is important to note that our current attacker models do not cover scenarios in which an adversary uses information from known users to infer sensitive data about a new or previously unseen user. In such cases, although the focus of this thesis is on identity-related features, an attacker who knows that the data belongs to a new user could potentially infer additional attributes, such as gender, age, or health conditions. These inferred attributes could then be used for cross-validation or combined with background knowledge, ultimately allowing the attacker to deduce the real identity of the new user. This represents a potential gap in our current framework, as our existing models do not account for the ability to infer additional attributes and ultimately deduce the real identity of a new user. Therefore, future work should first evaluate the feasibility and likelihood of such inference attacks. If

these attacks prove possible or likely to succeed, the privacy-preserving feature selection framework will need to be further refined to effectively mitigate this risk and ensure the protection of sensitive information in comprehensive service systems.

Legal Protection: Institutional Support for the Privacy of People with Disabilities

As discussed in the Introduction, the legal system plays an irreplaceable role in protecting privacy in comprehensive services for people with disabilities. Although technical measures can reduce the risk of data breaches to a certain extent, privacy protection without legal support is prone to becoming a mere formality in practical applications, especially in scenarios involving cross-platform and cross-institutional data sharing. According to our survey, there are significant regional differences in the legal protections of health data. For example, in the United States, the Health Insurance Portability and Accountability Act provides clear regulations for the collection, storage, and use of medical and health-related data. However, in most countries, specialized laws targeting health data (especially for vulnerable groups such as people with disabilities) remain relatively scarce, and legal frameworks are often fragmented and lack specificity. This gap not only increases the potential risk of privacy leakage, but also hinders the standardized development of comprehensive services. Therefore, we call on governments to establish a more comprehensive legal and policy framework, including clear definitions of the scope of health data use, compliant regulation of cross-institutional data exchange, and mechanisms for the participation of vulnerable populations such as persons with disabilities in privacy legislation and oversight. Through institutionalized legal safeguards, more robust and sustainable privacy protection can be provided to people with disabilities beyond technical measures, while also laying the foundation for the sustainable development of assistive living services.

In summary, this thesis not only explores methodological design and security enhancement, but also provides practical support for advancing privacy protection technologies toward greater fairness and inclusiveness in real-world applications.

BIBLIOGRAPHY

[1] Bahar Farahani et al., « Towards collaborative intelligent IoT eHealth: From device to fog, and cloud », en, *in*: *Microprocess. Microsyst.* 72.*102938* (2020), p. 102938.

[2] *Le handicap en chiffres*, [Online]. Available: https://drees.solidarites-sante.gouv.fr/publications-communique-de-presse-documents-de-reference/panoramas-de-la-drees/241128_Panorama_Handicap2024. Accessed: Juin.2025.

[3] World Health Organization, *Global report on health equity for persons with disabilities*, https://www.who.int/publications/i/item/9789240063600, Accessed: 2025-06-03, 2025.

[4] Mélanie Gréaux et al., « Health equity for persons with disabilities: a global scoping review on barriers and interventions in healthcare services », *in*: *International Journal for Equity in Health* 22.*1* (2023), p. 236.

[5] *General Data Protection Regulation*, [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj. Accessed: Jan.2024.

[6] *The European Health Data Space*, [Online]. Available: https://www.european-health-data-space.com/. Accessed: Jul.2025.

[7] Rogier Creemers and Graham Webster, *Translation: Personal information protection law of the people's republic of China*, [Online]. Available: https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/. Accessed: Jan.2024.

[8] David Stauss, *State data privacy legislation: Takeaways from 2022 and what to expect in 2023*, [Online]. Available: https://iapp.org/news/a/state-data-privacy-legislation-takeaways-from-2022-and-what-to-expect-in-2023/. Accessed: Jan.2024.

[9] *The Health Insurance Portability and Accountability Act*, [Online]. Available: https://www.hhs.gov/programs/hipaa/index.html. Accessed: Jul.2025.

[10] *India proposes Digital Personal Data Protection Act 2022*, [Online]. Available: https://iapp.org/news/a/india-proposes-digital-personal-data-protection-act-2022/. Accessed: Jan.2024.

[11] Dan Zhu, Xiao-Bai Li, and Shuning Wu, « Identity disclosure protection: A data reconstruction approach for privacy-preserving data mining », en, *in*: *Decis. Support Syst.* 48.*1* (2009), pp. 133–140.

[12] Stan Matwin et al., « A review of attribute disclosure control », *in*: *Studies in Computational Intelligence*, Cham: Springer, 2015, pp. 41–61.

[13] Iyiola E Olatunji et al., « A review of anonymization for healthcare data », en, *in*: *Big Data* (2022).

[14] David J Martin et al., « Worst-case background knowledge for privacy-preserving data publishing », *in*: *Proc. IEEE 23rd Int. Conf. Data Eng.* 2007, pp. 126–135.

[15] Khaled El Emam, Lucy Mosquera, and Xi Fang, « Validating a membership disclosure metric for synthetic health data », en, *in*: *J. Am. Med. Inform. Assoc.* 5.*4* (2022), ooac083.

[16] Raphael Spreitzer et al., « Systematic classification of side-channel attacks: A case study for mobile devices », *in*: *IEEE Commun. Surv. Tutor.* 20.*1* (2017), pp. 465–488.

[17] Narendra Kumar et al., « Health Care DNS Tunnelling Detection Method via Spiking Neural Network », *in*: *International Conference on Emergent Converging Technologies and Biomedical Systems*, Springer, 2023, pp. 715–725.

[18] Pardeep Kumar and Hoon-Jae Lee, « Security issues in healthcare applications using wireless medical sensor networks: A survey », *in*: *sensors* 12.*1* (2011), pp. 55–91.

[19] Meng Zhang, Anand Raghunathan, and Niraj K Jha, « Trustworthiness of medical devices and body area networks », *in*: *Proceedings of the IEEE* 102.*8* (2014), pp. 1174–1188.

[20] Peng Cheng et al., « SonarSnoop: active acoustic side-channel attacks », en, *in*: *Int. J. Inf. Secur.* 19.*2* (2020), pp. 213–228.

[21] Saurav Maji, Utsav Banerjee, and Anantha P Chandrakasan, « Leaky nets: Recovering embedded neural network models and inputs through simple power and timing side-channels—Attacks and defenses », *in*: *IEEE Internet of Things Journal* 8.*15* (2021), pp. 12079–12092.

[22] Yuan Cao et al., « A compact and low power RO PUF with high resilience to the EM side-channel attack and the SVM modelling attack of wireless sensor networks », *in*: *Sensors* 18.*2* (2018), p. 322.

[23] Fürkan Elibol, Uğur Sarac, and Işin Erer, « Realistic eavesdropping attacks on computer displays with low-cost and mobile receiver system », *in*: *Proc. Eur. Signal Process. Conf.* 2012, pp. 1767–1771.

[24] Michael Backes et al., « Acoustic side-channel attacks on printers », *in*: *Proc. USENIX Secur. Symp.* Vol. 9, 2010, pp. 307–322.

[25] Yonatan Gizachew Achamyeleh et al., « A Fly on the Wall-Exploiting Acoustic Side-Channels in Differential Pressure Sensors », *in*: *2024 Annual Computer Security Applications Conference (ACSAC)*, IEEE, 2024, pp. 240–256.

[26] Nan Li et al., « A review of security issues and solutions for precision health in Internet-of-Medical-Things systems », *in*: *Security and Safety* 2 (2023), p. 2022010.

[27] Abdul Rehman, Imran Razzak, and Guandong Xu, « Federated learning for privacy preservation of healthcare data from smartphone-based side-channel attacks », *in*: *IEEE Journal of Biomedical and Health Informatics* 27.*2* (2022), pp. 684–690.

[28] Raphael Spreitzer et al., « Exploiting data-usage statistics for website fingerprinting attacks on android », *in*: *Proc. 9th ACM Conf. Secur. Priv. Wirel. Mob. Netw.* 2016, pp. 49–60.

[29] Saurav Maji, Kyungmi Lee, and Anantha P Chandrakasan, « SparseLeakyNets: Classification prediction attack over sparsity-aware embedded neural networks using timing side-channel information », *in*: *IEEE Computer Architecture Letters* (2024).

[30] Mehwash Weqar, Shabana Mehfuz, and Dhawal Gupta, « DNS Traffic Monitoring to Access Vulnerability in the Internet of Healthcare Things Networks: A Survey », *in*: *2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON)*, IEEE, 2023, pp. 89–94.

[31] *DNS Flaws in Millions of IoT Devices Pose Remote Attack, Exfiltration Risk*, [Online]. Available: `https://www.techtarget.com/healthtechsecurity/answer/DNS-Flaws-in-Millions-of-IoT-Devices-Pose-Remote-Attack-Exfiltration-Risk`. Accessed: Sep.2025.

[32] Marta Moure-Garrido, Carlos Garcia-Rubio, and Celeste Campo, « Reducing DNS Traffic to Enhance Home IoT Device Privacy », *in*: *Sensors* 24.*9* (2024), p. 2690.

[33] Alexandru Vulpe et al., « Enabling security services in socially assistive robot scenarios for healthcare applications », *in*: *Sensors* 21.*20* (2021), p. 6912.

[34] Ghazale Amel Zendehdel et al., « Automated security assessment framework for wearable ble-enabled health monitoring devices », *in*: *ACM Transactions on Internet Technology (TOIT)* 22.*1* (2021), pp. 1–31.

[35] Prithwi Bagchi et al., « Post quantum lattice-based secure framework using aggregate signature for ambient intelligence assisted blockchain-based IoT applications », *in*: *IEEE Internet of Things Magazine* 6.*1* (2023), pp. 52–58.

[36] Xinyao Zheng et al., « Inputsnatch: Stealing input in llm services via timing side-channel attacks », *in*: *arXiv preprint arXiv:2411.18191* (2024).

[37] M_Ozgur Ozmen, Habiba Farrukh, and ZBerkay Celik, « Physical Side-Channel Attacks against Intermittent Devices », *in*: (2024).

[38] Shuo Chen et al., « Side-channel leaks in web applications: A reality today, a challenge tomorrow », *in*: *2010 IEEE Symposium on Security and Privacy*, IEEE, 2010, pp. 191–206.

[39] Sandra Siby et al., « Encrypted DNS → privacy? A traffic analysis perspective », *in*: (2019), arXiv:1906.09682.

[40] Taha Belkhouja et al., « Light-weight solution to defend implantable medical devices against man-in-the-middle attack », *in*: *2018 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2018, pp. 1–5.

[41] Tal Melamed, « An active man-in-the-middle attack on bluetooth smart devices », *in*: *Safety and Security Studies* 15 (2018), p. 2018.

[42] Muhammed Ali Yurdagul and Husrev Taha Sencar, « Blekeeper: Response time behavior based man-in-the-middle attack detection », *in*: *2021 IEEE Security and Privacy Workshops (SPW)*, IEEE, 2021, pp. 214–220.

[43] Osman Salem et al., « Man-in-the-middle attack mitigation in internet of medical things », *in*: *IEEE Transactions on Industrial Informatics* 18.*3* (2021), pp. 2053–2062.

[44] Zhibo Wang et al., « Beyond inferring class representatives: User-level privacy leakage from federated learning », *in*: *Proc. IEEE Conf. Comput. Commun.* 2019, pp. 2512–2520.

[45] Reza Shokri et al., « Membership inference attacks against machine learning models », *in*: *Proc. IEEE Symp. Secur. Priv.* 2017, pp. 3–18.

[46] Ziqi Zhang, Chao Yan, and Bradley A Malin, « Membership inference attacks against synthetic health data », *in*: *Journal of biomedical informatics* 125 (2022), p. 103977.

[47] Gaoyang Liu et al., « SocInf: Membership inference attacks on social media health data with machine learning », *in*: *IEEE Transactions on Computational Social Systems* 6.*5* (2019), pp. 907–921.

[48] Yuheng Zhang et al., « The secret revealer: Generative model-inversion attacks against deep neural networks », *in*: *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 2020, pp. 253–261.

[49] Asfandyar Azhar, Paul Thielen, and Curtis Langlotz, « MEDUSA: Medical Data Under Shadow Attacks via Hybrid Model Inversion », *in*: *International Conference on Artificial Intelligence and Statistics*, PMLR, 2025, pp. 1378–1386.

[50] Moomal Bukhari et al., « A Representation Learning Induced Property Inference Attack on Machine Learning Models for E-Health », *in*: *GLOBECOM 2024-2024 IEEE Global Communications Conference*, IEEE, 2024, pp. 2828–2833.

[51] Nicholas Carlini et al., « The secret sharer: Evaluating and testing unintended memorization in neural networks », *in*: *Proc. 28th USENIX Secur. Symp.* 2019, pp. 267–284.

[52] Edoardo Debenedetti et al., « Privacy side channels in machine learning systems », *in*: *33rd USENIX Security Symposium (USENIX Security 24)*, 2024, pp. 6861–6848.

[53] Kaiyang Li et al., « Adversarial privacy-preserving graph embedding against inference attack », *in*: *IEEE Internet Things J.* 8.*8* (2020), pp. 6904–6915.

[54] Jian Wang et al., « A survey on privacy preserving data mining », *in*: *Proc. IEEE 1st Int. Workshop Database Technol. Appl.* 2009, pp. 111–114.

[55] Bing Su et al., « K-anonymity privacy protection algorithm for multi-dimensional data against skewness and similarity attacks », en, *in*: *Sensors* 23.*3* (2023), p. 1554.

[56] Srivatsava Ranjit Ganta, Shiva Prasad Kasiviswanathan, and Adam Smith, « Composition attacks and auxiliary information in data privacy », *in*: *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2008, pp. 265–273.

[57] *Cerballiance a constaté un accès non autorisé à vos données*, [Online]. Available: `https://www.clubic.com/actualite-558999-cerballiance-a-constate-un-acces-non-autorise-a-vos-donnees-non-ce-sms-n-est-pas-du-phishing.html`. Accessed: Sep.2025.

[58] Billy Bob Brumley and Nicola Tuveri, « Remote timing attacks are still practical », *in*: *Computer Security*, Berlin, Germany: Springer, 2011, pp. 355–371.

[59] Pepe Vila and Boris Köpf, « Loophole: Timing attacks on shared event loops in chrome », *in*: *Proc. 26th USENIX Secur. Symp.* 2017, pp. 849–864.

[60] Yuan Zhang et al., « Towards thwarting template side-channel attacks in secure cloud deduplications », *in*: *IEEE Transactions on Dependable and Secure Computing* 18.*3* (2019), pp. 1008–1018.

[61] Stjepan Picek et al., « SoK: Deep learning-based physical side-channel analysis », en, *in*: *ACM Comput. Surv.* 55.*11* (2023), pp. 1–35.

[62] Vincent Bindschaedler et al., « The tao of inference in privacy-protected databases », en, *in*: *Proc. VLDB Endow.* 11.*11* (2018), pp. 1715–1728.

[63] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian, « T-closeness: Privacy beyond k-anonymity and l-diversity », *in*: *Proc. IEEE 23rd Int. Conf. Data Eng.* 2007, pp. 106–115.

[64] J Andrew, R Jennifer Eunice, and J Karthikeyan, « An anonymization-based privacy-preserving data collection protocol for digital health data », *in*: *Frontiers in public health* 11 (2023), p. 1125011.

[65] Shiva Prasad Kasiviswanathan et al., « What can we learn privately? », *in*: *SIAM J. Comput.* 40.*3* (2011), pp. 793–826.

[66] Mengmeng Yang et al., « Local differential privacy and its applications: A comprehensive survey », *in*: (2020), arXiv:2008.03686.

[67] Andrea Bittau et al., « Prochlo: Strong privacy for analytics in the crowd », *in*: *Proc. 26th ACM Symp. Operating Syst. Princ.* 2017, pp. 441–459.

[68] Venkatesan T Chakaravarthy et al., « Efficient techniques for document sanitization », *in*: *Proc. 17th ACM Conf. Inf. Knowl. Manage.* 2008, pp. 843–852.

[69] David Sánchez, Montserrat Batet, and Alexandre Viejo, « Utility-preserving privacy protection of textual healthcare documents », *in*: *Journal of biomedical informatics* 52 (2014), pp. 189–198.

[70] Balamurugan Anandan et al., « T-Plausibility: Generalizing words to desensitize text », en, *in*: *Trans. Data Priv.* 5.*3* (2012), pp. 505–534.

[71] David Sánchez and Montserrat Batet, « C-sanitized: A privacy model for document redaction and sanitization », en, *in*: *J. Assoc. Inf. Sci. Technol.* 67.*1* (2016), pp. 148–163.

[72] Fadi Hassan, David Sanchez, and Josep Domingo-Ferrer, « Utility-preserving privacy protection of textual documents via word embeddings », *in*: *IEEE Trans. Knowl. Data Eng.* 35.*1* (2021), pp. 1058–1071.

[73] Francine Chen, John Adcock, and Shruti Krishnagiri, « Audio privacy: Reducing speech intelligibility while preserving environmental sounds », *in*: *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 733–736.

[74] Eric C Larson et al., « Accurate and privacy preserving cough sensing using a low-cost microphone », *in*: *Proc. 13th Int. Conf. Ubiquitous comput.* 2011, pp. 375–384.

[75] Sumeet Kumar et al., « Sound shredding: Privacy preserved audio sensing », *in*: *Proc. 16th Int. Workshop Mob. Comput. Syst. Appl.* 2015, pp. 135–140.

[76] Xiaosong Ma et al., « You can hear but you cannot record: Privacy protection by jamming audio recording », *in*: *Proc. IEEE Int. Conf. Commun.* 2021, pp. 1–6.

[77] Jaemin Lim et al., « Overo: Sharing Private Audio Recordings », *in*: *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* 2022, pp. 1933–1946.

[78] Wei-Cheng Wang et al., « An opt-in framework for privacy protection in audio-based applications », *in*: *IEEE Pervasive Comput.* 21.*4* (2022), pp. 17–24.

[79] Xiaorui Zhang et al., « Robust reversible audio watermarking scheme for telemedicine and privacy protection », en, *in*: *Comput. Mater. Contin.* 71.*2* (2022), pp. 3035–3050.

[80] Zhihua Xia et al., « A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing », *in*: *IEEE Trans. Inf. Forensics Secur.* 11.*11* (2016), pp. 2594–2608.

[81] Meng Shen et al., « Content-based multi-source encrypted image retrieval in clouds with privacy preservation », en, *in*: *Future Gener. Comput. Syst.* 109 (2020), pp. 621–632.

[82] Fei Yu et al., « Privacy protection of medical data based on multi-scroll memristive Hopfield neural network », *in*: *IEEE Transactions on Network Science and Engineering* 10.*2* (2022), pp. 845–858.

[83] Denghui Zhang et al., « A privacy protection framework for medical image security without key dependency based on visual cryptography and trusted computing », *in*: *Computational Intelligence and Neuroscience* 2023.*1* (2023), p. 6758406.

[84] Liyue Fan, « Image Pixelization with Differential Privacy », *in*: *Data and Applications Security and Privacy*, Cham: Springer, 2018, pp. 148–162.

[85] Hisaichi Shibata et al., « Practical medical image generation with provable privacy protection based on denoising diffusion probabilistic models for high-resolution volumetric images », *in*: *Applied Sciences* 14.*8* (2024), p. 3489.

[86] Jun Yu et al., « IPrivacy: Image privacy protection by identifying sensitive objects via deep multi-task learning », *in*: *IEEE Trans. Inf. Forensics Secur.* 12.*5* (2016), pp. 1005–1016.

[87] Jinao Yu et al., « GAN-based differential private image privacy protection framework for the Internet of Multimedia Things », en, *in*: *Sensors* 21.*1* (2020), p. 58.

[88] Wenqi Li et al., « Privacy-preserving federated brain tumour segmentation », *in*: *Machine Learning in Medical Imaging*, Cham: Springer, 2019, pp. 133–141.

[89] Zhenzhong Kuang et al., « Unnoticeable synthetic face replacement for image privacy protection », en, *in*: *Neurocomputing* 457 (2021), pp. 322–333.

[90] Jingyi Cao et al., « Hiding among your neighbors: Face image privacy protection with differential private k-anonymity », *in*: *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.* 2022, pp. 1–6.

[91] Frédéric Dufaux and Touradj Ebrahimi, « A framework for the validation of privacy protection solutions in video surveillance », *in*: *Proc. IEEE Int. Conf. Multimedia Expo*, 2010, pp. 66–71.

[92] Khalid M Hosny et al., « Privacy Protection in Surveillance Videos Using Block Scrambling-Based Encryption and DCNN-Based Face Detection », *in*: *IEEE Access* 10 (2022), pp. 106750–106769.

[93] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo, « Learning to anonymize faces for privacy preserving action detection », *in*: *Computer Vision*, Cham: Springer, 2018, pp. 639–655.

[94] Zhaodong Sun and Xiaobai Li, « Privacy-phys: Facial video-based physiological modification for privacy protection », *in*: *IEEE Signal Processing Letters* 29 (2022), pp. 1507–1511.

[95] Jixin Liu et al., « Combined CNN/RNN video privacy protection evaluation method for monitoring home scene violence », en, *in*: *Comput. Electr. Eng.* 106.*108614* (2023), p. 108614.

[96] Honghui Xu et al., « Audio-visual autoencoding for privacy-preserving video streaming », *in*: *IEEE Internet Things J.* 9.*3* (2021), pp. 1749–1761.

[97] Yuying Qiu et al., « A novel generative model for face privacy protection in video surveillance with utility maintenance », en, *in*: *Appl. Sci.* 12.*14* (2022), p. 6962.

[98] Haohua Du et al., « PatronuS: A system for privacy-preserving cloud video surveillance », *in*: *IEEE J. Sel. Areas Commun.* 38.*6* (2020), pp. 1252–1261.

[99] Han Wang et al., « Publishing video data with indistinguishable objects », en, *in*: *Adv. Database Technol.* 2020 (2020), pp. 323–334.

[100] Bingquan Zhu et al., « Deepfakes for medical video DE-identification: Privacy protection and diagnostic information preservation », *in*: *Proc. AAAI/ACM Conf. AI Ethics Soc.* 2020, pp. 414–420.

[101] Alem Fitwi et al., « Privacy-preserving Surveillance as an edge service based on lightweight video protection schemes using face DE-identification and window masking », en, *in*: *Electronics* 10.*3* (2021), p. 236.

[102] E Shi et al., « Privacy-preserving aggregation of time-series data », en, *in*: *Proc. Annu. Netw. Distrib. Syst. Secur. Symp.* 2011.

[103] Marc Joye and Benoıt Libert, « A scalable scheme for privacy-preserving aggregation of time-series data », *in*: *Financial Cryptography and Data Security*, Berlin, Germany: Springer, 2013, pp. 111–125.

[104] Lingjuan Lyu et al., « Privacy-preserving collaborative deep learning with application to human activity recognition », *in*: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 1219–1228.

[105] Mohammad Malekzadeh, Richard G Clegg, and Hamed Haddadi, « Replacement autoencoder: A privacy-preserving algorithm for sensory data analysis », *in*: *arXiv preprint arXiv:1710.06564* (2017).

[106] Mohammad Malekzadeh et al., « Protecting sensory data against sensitive inferences », *in*: *Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems*, 2018, pp. 1–6.

[107] Leonardo Bigelli et al., « Privacy preservation in sensor-based Human Activity Recognition through autoencoders for low-power IoT devices », *in*: *Internet of Things* 26 (2024), p. 101189.

[108] Zhiwen Xiao et al., « A federated learning system with enhanced feature extraction for human activity recognition », *in*: *Knowledge-Based Systems* 229 (2021), p. 107338.

[109] Kongyang Chen et al., « Federated unlearning for human activity recognition », *in*: *arXiv preprint arXiv:2404.03659* (2024).

[110] Longxiang Shi et al., « Semantic health knowledge graph: Semantic integration of heterogeneous medical knowledge and services », en, *in*: *Biomed Res. Int.* 2017 (2017), p. 2858423.

[111] Kun Liu and Evimaria Terzi, « Towards identity anonymization on graphs », *in*: *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2008, pp. 93–106.

[112] Lei Zou, Lei Chen, and M Tamer Özsu, « K-automorphism: A general framework for privacy preserving network publication », en, *in*: *Proc. VLDB Endow.* 2.*1* (2009), pp. 946–957.

[113] M E Skarkala et al., « Privacy preservation by k-anonymization of weighted social networks », *in*: *Proc. IEEE/ACM Int. Conf. Adv. Soc. Netw. Anal. Min.* 2012, pp. 423–428.

[114] Michael Hay et al., « Resisting structural re-identification in anonymized social networks », en, *in*: *Proc. VLDB Endow.* 1.*1* (2008), pp. 102–114.

[115] Alessandra Sala et al., « Sharing graphs using differentially private graph models », *in*: *Proc. ACM SIGCOMM Conf. Internet Meas. Conf.* 2011, pp. 81–98.

[116] Elena Zheleva and Lise Getoor, « Preserving the privacy of sensitive relationships in graph data », *in*: *Privacy, Security, and Trust in KDD*, Berlin, Germany: Springer, 2008, pp. 153–171.

[117] Prateek Mittal, Charalampos Papamanthou, and Dawn Song, « Preserving link privacy in social network based systems », *in*: (2012), arXiv:1208.6189.

[118] Mohamed Elhoseny et al., « IoT solution for AI-enabled PRIVACY-PREServing with big data transferring: an application for healthcare using blockchain », *in*: *Energies* 14.*17* (2021), p. 5364.

[119] Robin C Geyer, Tassilo Klein, and Moin Nabi, « Differentially private federated learning: A client level perspective », *in*: (2017), arXiv:1712.07557.

[120] Pathum Chamikara Mahawaga Arachchige et al., « Local differential privacy for deep learning », *in*: *IEEE Internet Things J.* 7.*7* (2019), pp. 5827–5842.

[121] Chugui Xu et al., « EdgeSanitizer: Locally differentially private deep inference at the edge for mobile data analytics », *in*: *IEEE Internet Things J.* 6.*3* (2019), pp. 5140–5151.

[122] Soroosh Tayebi Arasteh et al., « Private, fair and accurate: Training large-scale, privacy-preserving ai models in medical imaging », *in*: *arXiv preprint arXiv:2302.01622* (2023).

[123] Soroosh Tayebi Arasteh et al., « Securing collaborative medical AI by using differential privacy: Domain transfer for classification of chest radiographs », *in*: *Radiology: Artificial Intelligence* 6.*1* (2023), e230212.

[124] Kun Liu, H Kargupta, and J Ryan, « Random projection-based multiplicative data perturbation for privacy preserving distributed data mining », *in*: *IEEE Trans. Knowl. Data Eng.* 18.*1* (2006), pp. 92–106.

[125] S Y Kung, « Compressive privacy: From information\/estimation theory to machine learning [lecture notes] », *in*: *IEEE Signal Process. Mag.* 34.*1* (2017), pp. 94–112.

[126] Reza Shokri and Vitaly Shmatikov, « Privacy-preserving deep learning », *in*: *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.* 2015, pp. 1310–1321.

[127] Li Zhang et al., « Homomorphic encryption-based privacy-preserving federated learning in IoT-enabled healthcare system », *in*: *IEEE Transactions on Network Science and Engineering* 10.*5* (2022), pp. 2864–2880.

[128] Rajesh Kumar et al., « Blockchain and homomorphic encryption based privacy-preserving model aggregation for medical images », *in*: *Computerized Medical Imaging and Graphics* 102 (2022), p. 102139.

[129] Muhammad Asad et al., « SHFL: K-anonymity-based secure hierarchical federated learning framework for smart healthcare systems », en, *in*: *Future Internet* 14.*11* (2022), p. 338.

[130] Konstantinos Kalodanis et al., « A Privacy-Preserving and Attack-Aware AI Approach for High-Risk Healthcare Systems Under the EU AI Act », *in*: *Electronics* 14.*7* (2025), p. 1385.

[131] Nicolas Papernot et al., « Semi-supervised knowledge transfer for deep learning from private training data », *in*: (2016), arXiv:1610.05755.

[132] Virat Shejwalkar and Amir Houmansadr, « Membership privacy for machine learning models through knowledge transfer », *in*: *Proc. Conf. AAAI Artif. Intell.* 35.*11* (2021), pp. 9549–9557.

[133] Xinghui Fei et al., « A Systematic Study on the Privacy Protection Mechanism of Natural Language Processing in Medical Health Records », *in*: *2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE)*, IEEE, 2024, pp. 1819–1824.

[134] Nazish Khalid et al., « Privacy-preserving artificial intelligence in healthcare: Techniques and applications », *in*: *Computers in Biology and Medicine* 158 (2023), p. 106848.

[135] Steven M Williamson and Victor Prybutok, « Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare », *in*: *Applied Sciences* 14.*2* (2024), p. 675.

[136] Wang Ren et al., « Privacy-preserving using homomorphic encryption in Mobile IoT systems », en, *in*: *Comput. Commun.* 165 (2021), pp. 105–111.

[137] Vishal A Thakor, Mohammad Abdur Razzaque, and Muhammad R A Khandaker, « Lightweight cryptography algorithms for resource-constrained IoT devices: A review, comparison and research opportunities », *in*: *IEEE Access* 9 (2021), pp. 28177–28193.

[138] Haqi Khalid et al., « Cross-SN: a lightweight authentication scheme for a multi-server platform using IoT-based wireless medical sensor network », *in*: *Electronics* 10.*7* (2021), p. 790.

[139] Qi Feng et al., « A survey on privacy protection in blockchain system », en, *in*: *J. Netw. Comput. Appl.* 126 (2019), pp. 45–58.

[140] Shuyun Shi et al., « Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey », en, *in*: *Comput. Secur.* 97.*101966* (2020), p. 101966.

[141] Thomas Popp, Stefan Mangard, and Elisabeth Oswald, « Power analysis attacks and countermeasures », *in*: *IEEE Des. Test Comput.* 24.*6* (2007), pp. 535–543.

[142] Christophe Clavier, Jean-Sébastien Coron, and Nora Dabbous, « Differential power analysis in the presence of hardware countermeasures », *in*: *Cryptographic Hardware and Embedded Systems*, Berlin, Germany: Springer, 2000, pp. 252–263.

[143] Jean-Jacques Quisquater and David Samyde, « ElectroMagnetic analysis (EMA): Measures and counter-measures for smart cards », *in*: *Smart Card Programming and Security*, Berlin, Germany: Springer, 2001, pp. 200–210.

[144] Ken Mai, « Side channel attacks and countermeasures », *in*: *Introduction to Hardware Security and Trust*, Springer, 2011, pp. 175–194.

[145] Tromer Eran, « Acoustic cryptanalysis: on nosy people and noisy machines », en, *in*: *Eurocrypt Rump Session* (2004).

[146] Kanthakumar Pongaliur et al., « Securing sensor nodes against side channel attacks », *in*: *Proc. 11th IEEE High Assur. Syst. Eng. Symp.* 2008, pp. 353–361.

[147] Chuhan Gao et al., « Privacy protection for audio sensing against multi-microphone adversaries », *in*: *Proceedings on Privacy Enhancing Technologies* (2019).

[148] Jingchao Sun et al., « VISIBLE: Video-assisted keystroke inference from tablet backside motion », *in*: *Proc. Netw. Distrib. Syst. Secur. Symp.* 2016.

[149] Prakash Shrestha, Manar Mohamed, and Nitesh Saxena, « Slogger: Smashing motion-based touchstroke logging with transparent system noise », *in*: *Proc. 9th ACM Conf. Secur. Priv. Wirel. Mob. Netw.* 2016, pp. 67–77.

[150] Qiuyu Xiao, Michael K Reiter, and Yinqian Zhang, « Mitigating storage side channels using statistical privacy mechanisms », *in*: *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.* 2015, pp. 1582–1594.

[151] Emmanuel Owusu et al., « ACCessory: Password inference using accelerometers on smartphones », *in*: *Proc. 12th Workshop Mob. Comput. Syst. Appl.* 2012, pp. 1–6.

[152] Nan Zhang et al., « Leave me alone: App-level protection against runtime information gathering on android », *in*: *Proc. IEEE Symp. Secur. Priv.* 2015, pp. 915–930.

[153] Cynthia Dwork et al., « Our data, ourselves: Privacy via distributed noise generation », *in*: *Advances in Cryptology*, Berlin, Germany: Springer, 2006, pp. 486–503.

[154] Abbas Acar et al., « Achieving secure and differentially private computations in multiparty settings », *in*: *2017 IEEE Symposium on Privacy-Aware Computing (PAC)*, IEEE, 2017, pp. 49–59.

[155] Yi Yang et al., « Towards statistically strong source anonymity for sensor networks », en, *in*: *ACM Trans. Sens. Netw.* 9.*3* (2013), pp. 1–23.

[156] Noah Apthorpe et al., « Keeping the smart home private with smart(er) IoT traffic shaping », *in*: (2018), arXiv:1812.00955.

[157] Nelson Prates et al., « A defense mechanism for timing-based side-channel attacks on IoT traffic », *in*: *Proc. IEEE Global Commun. Conf.* 2020, pp. 1–6.

[158] Noah Apthorpe et al., « Spying on the smart home: Privacy attacks and defenses on encrypted IoT traffic », *in*: (2017), arXiv:1708.05044.

[159] Charles V Wright, Scott E Coull, and Fabian Monrose, « Traffic Morphing: An Efficient Defense Against Statistical Traffic Analysis », en, *in*: *Proc. Netw. Distrib. Syst. Secur. Symp.* Vol. 9, 2009.

[160] Sijie Xiong, Anand D Sarwate, and Narayan B Mandayam, « Defending against packet-size side-channel attacks in iot networks », *in*: *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* 2018, pp. 2027–2031.

[161] Cynthia Dwork, « Differential Privacy », *in*: *Automata, Languages and Programming*, Berlin, Germany: Springer, 2006, pp. 1–12.

[162] Brendan Avent et al., « {BLENDER}: Enabling local search with a hybrid differential privacy model », *in*: *Proc. 26th USENIX Secur. Symp.* 2017, pp. 747–764.

[163] Wei Wang, Lei Chen, and Qian Zhang, « Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation », *in*: *Computer Networks* 88 (2015), pp. 136–148.

[164] Xiaojun Zhang et al., « CIPPPA: Conditional identity privacy-preserving public auditing for cloud-based WBANs against malicious auditors », *in*: *IEEE transactions on cloud Computing* 9.*4* (2019), pp. 1362–1375.

[165] Junqi Guo, Minghui Yang, and Boxin Wan, « A practical privacy-preserving publishing mechanism based on personalized k-anonymity and temporal differential privacy for wearable IoT applications », en, *in*: *Symmetry* 13.*6* (2021), p. 1043.

[166] Ankhbayar Otgonbayar et al., « K-VARP: K-anonymity for varied data streams via partitioning », en, *in*: *Inf. Sci.* 467 (2018), pp. 238–255.

[167] K Mohana Prabha and P Vidhya Saraswathi, « Suppressed K-Anonymity Multi-Factor Authentication Based Schmidt-Samoa Cryptography for privacy preserved data access in cloud computing », en, *in*: *Comput. Commun.* 158 (2020), pp. 85–94.

[168] Karuna Arava and Sumalatha Lingamgunta, « Adaptive k-anonymity approach for privacy preserving in cloud », en, *in*: *Arab. J. Sci. Eng.* 45.*4* (2020), pp. 2425–2432.

[169] Ashwin Machanavajjhala et al., « *L*-diversity: Privacy beyond *k*-anonymity », en, *in*: *ACM Trans. Knowl. Discov. Data* 1.*1* (2007), p. 3.

[170] Raymond Chi-Wing Wong et al., « (α, k)-anonymity: An enhanced k-anonymity model for privacy preserving data publishing », *in*: *Proc. 12th ACM SIGKDD Int Conf. Knowl. Discov. Data Min.* 2006, pp. 754–759.

[171] T M Truta and B Vinay, « Privacy protection: P-sensitive k-anonymity property », *in*: *Proc. IEEE 22nd Int. Conf. Data Eng. Workshops*, 2006, pp. 94–94.

[172] Tiancheng Li et al., « Slicing: A new approach for privacy preserving data publishing », *in*: *IEEE Trans. Knowl. Data Eng.* 24.*3* (2012), pp. 561–574.

[173] Mohammed BinJubier et al., « Slicing-based enhanced method for privacy-preserving in publishing big data », en, *in*: *Comput. Mater. Contin.* 72.*2* (2022), pp. 3665–3686.

[174] Marcel Keller, Emmanuela Orsini, and Peter Scholl, « MASCOT: Faster malicious arithmetic secure computation with oblivious transfer », *in*: *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* 2016, pp. 830–842.

[175] Parsa Sarosh et al., « Secret sharing-based personal health records management for the internet of health things », en, *in*: *Sustain. Cities Soc.* 74.*103129* (2021), p. 103129.

[176] Hanlin Zhang et al., « Cloud storage for electronic health records based on secret sharing with verifiable reconstruction outsourcing », *in*: *IEEE Access* 6 (2018), pp. 40713–40722.

[177] Michael O Rabin, « How to exchange secrets with oblivious transfer », en, *in*: *Cryptol. ePrint Arch.* (2005).

[178] Huijie Yang et al., « A privacy-preserving data transmission scheme based on oblivious transfer and blockchain technology in the smart healthcare », en, *in*: *Secur. Commun. Netw.* 2021 (2021), pp. 1–12.

[179] Benny Chor et al., « Private information retrieval », en, *in*: *J. ACM* 45.*6* (1998), pp. 965–981.

[180] Feng Chen et al., « PRECISE: PRivacy-preserving cloud-assisted quality improvement service in healthcare », *in*: *Proc. IEEE 8th Int. Conf. Syst. Biol.* 2014, pp. 176–183.

[181] Mauro Barni, Jorge Guajardo, and Riccardo Lazzeretti, « Privacy preserving evaluation of signal quality with application to ECG analysis », *in*: *Proc. IEEE Int. Workshop Inf. Forensics Secur.* 2010, pp. 1–6.

[182] Benny Pinkas, Thomas Schneider, and Michael Zohner, « Scalable private set intersection based on OT extension », en, *in*: *ACM Trans. Priv. Secur.* 21.*2* (2018), pp. 1–35.

[183] Yalian Qian et al., « Profile matching for IoMT: A verifiable private set intersection scheme », en, *in*: *IEEE J. Biomed. Health Inform.* 25.*10* (2021), pp. 3794–3803.

[184] Yi-Wei Wang and Ja-Ling Wu, « A privacy-preserving symptoms retrieval system with the aid of Homomorphic Encryption and Private Set Intersection schemes », en, *in*: *Algorithms* 16.*5* (2023), p. 244.

[185] Atsuko Miyaji, Kazuhisa Nakasho, and Shohei Nishida, « Privacy-preserving integration of medical data : A practical multiparty Private Set Intersection: A practical multiparty private set intersection », en, *in*: *J. Med. Syst.* 41.*3* (2017), p. 37.

[186] Milad Nasr, Reza Shokri, and Amir Houmansadr, « Machine learning with membership privacy using adversarial regularization », *in*: *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.* 2018, pp. 634–646.

[187] Jiacheng Li, Ninghui Li, and Bruno Ribeiro, « Membership inference attacks and defenses in classification models », *in*: *Proc. 11th ACM Conf. Data Appl. Secur. Priv.* 2021, pp. 5–16.

[188] Raphael Bost et al., « Machine learning classification over encrypted data », en, *in*: *IACR Cryptol. ePrint Arch.* (2014).

[189] Ran Gilad-Bachrach et al., « CryptoNets: Applying neural networks to encrypted data with high throughput and accuracy », *in*: *Proc. 33rd Int. Conf. Mach. Learn.* Vol. 48, 2016, pp. 201–210.

[190] Ruyan Wang et al., « Cloud-based secure human action recognition with fully homomorphic encryption », *in*: *The Journal of Supercomputing* 81.*1* (2025), pp. 1–27.

[191] Michael Godfrey and Mohammad Zulkernine, « Preventing cache-based side-channel attacks in a cloud environment », *in*: *IEEE Trans. Cloud Comput.* 2.*4* (2014), pp. 395–408.

[192] Fangfei Liu et al., « CATalyst: Defeating last-level cache side channel attacks in cloud computing », *in*: *Proc. IEEE Int. Symp. High Perform. Comput. Archit.* 2016, pp. 406–418.

[193] Tianwei Zhang, Yinqian Zhang, and Ruby B Lee, « CloudRadar: A real-time side-channel attack detection system in clouds », *in*: *Research in Attacks, Intrusions, and Defenses*, Cham: Springer, 2016, pp. 118–140.

[194] Aastha Mehta et al., « Pacer: Comprehensive Network {Side-Channel} Mitigation in the Cloud », *in*: *31st USENIX Security Symposium (USENIX Security 22)*, 2022, pp. 2819–2838.

[195] Katsuhiko Kaji et al., « HASC-IPSC: indoor pedestrian sensing corpus with a balance of gender and age for indoor positioning and floor-plan generation researches », *in*: *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, 2013, pp. 605–610.

[196] Gary M Weiss, « Wisdm smartphone and smartwatch activity and biometrics dataset », *in*: *UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set* 7.*133190-133202* (2019), p. 5.

[197] Eftim Zdravevski et al., « Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering », *in*: *Ieee Access* 5 (2017), pp. 5262–5280.

[198] Nadeem Ahmed, Jahir Ibna Rafiq, and Md Rashedul Islam, « Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model », *in*: *Sensors* 20.*1* (2020), p. 317.

[199] Rung-Ching Chen et al., « Selecting critical features for data classification based on machine learning methods », *in*: *Journal of Big Data* 7.*1* (2020), p. 52.

[200] Seham Hussein, Michael Barlow, and Erandi Lakshika, « Activity-independent person identification based on daily activities using wearable sensors », *in*: *AI 2020: Advances in Artificial Intelligence: 33rd Australasian Joint Conference, AI 2020, Canberra, ACT, Australia, November 29–30, 2020, Proceedings 33*, Springer, 2020, pp. 383–394.

[201] Fei Luo et al., « Activity-based person identification using multimodal wearable sensor data », *in*: *IEEE Internet of Things Journal* 10.*2* (2022), pp. 1711–1723.

[202] Le Ngu Nguyen et al., « Multi-objective Feature Selection in Remote Health Monitoring Applications », *in*: *arXiv preprint arXiv:2401.05538* (2024).

[203] Angelica Poli et al., « Balancing activity recognition and privacy preservation with a multi-objective evolutionary algorithm », *in*: *International Conference on Smart Objects and Technologies for Social Good*, Springer, 2021, pp. 3–17.

[204] Eckart Zitzler, Marco Laumanns, and Stefan Bleuler, « A tutorial on evolutionary multiobjective optimization », *in*: *Metaheuristics for multiobjective optimisation* (2004), pp. 3–37.

[205] Kalyanmoy Deb and Himanshu Jain, « An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints », *in*: *IEEE transactions on evolutionary computation* 18.*4* (2013), pp. 577–601.

[206] Kalyanmoy Deb et al., « A fast and elitist multiobjective genetic algorithm: NSGA-II », *in*: *IEEE transactions on evolutionary computation* 6.*2* (2002), pp. 182–197.

[207] Julian Blank and Kalyanmoy Deb, « Pymoo: Multi-objective optimization in python », *in*: *Ieee access* 8 (2020), pp. 89497–89509.

[208] Pau Climent-Pérez and Francisco Florez-Revuelta, « Privacy-preserving human action recognition with a many-objective evolutionary algorithm », *in*: *Sensors* 22.*3* (2022), p. 764.

[209] Leo Breiman, « Random forests », *in*: *Machine learning* 45 (2001), pp. 5–32.

[210] Dipanwita Thakur and Suparna Biswas, « An integration of feature extraction and guided regularized random forest feature selection for smartphone based human activity recognition », *in*: *Journal of Network and Computer Applications* 204 (2022), p. 103417.

[211] Wahyu Caesarendra and Tegoeh Tjahjowidodo, « A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing », *in*: *Machines* 5.*4* (2017), p. 21.

[212] Chelsea Dobbins, Reza Rawassizadeh, and Elaheh Momeni, « Detecting physical activity within lifelogs towards preventing obesity and aiding ambient assisted living », *in*: *Neurocomputing* 230 (2017), pp. 110–132.

[213] Wesllen Sousa Lima et al., « Human activity recognition using inertial sensors in a smartphone: An overview », *in*: *Sensors* 19.*14* (2019), p. 3213.

[214] Marília Barandas et al., « TSFEL: Time series feature extraction library », *in*: *SoftwareX* 11 (2020), p. 100456.

[215] Charles R Harris et al., « Array programming with NumPy », *in*: *Nature* 585.*7825* (2020), pp. 357–362.

[216] Dipanwita Thakur and Suparna Biswas, « Permutation importance based modified guided regularized random forest in human activity recognition with smartphone », *in*: *Engineering Applications of Artificial Intelligence* 129 (2024), p. 107681.

[217] Zhibo Wang et al., « Towards pattern-aware privacy-preserving real-time data collection », *in*: *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, IEEE, 2020, pp. 109–118.

[218] Lingjuan Lyu et al., « Privacy and robustness in federated learning: Attacks and defenses », *in*: *IEEE transactions on neural networks and learning systems* 35.*7* (2022), pp. 8726–8746.

[219] Jiapeng Wu et al., « Augment then Smooth: Reconciling Differential Privacy with Certified Robustness », *in*: *Transactions on Machine Learning Research* (2024), URL: `https://openreview.net/forum?id=YN0IcnXqsr`.

[220] Li Qiyang et al., « Multi-objective feature selection for user privacy protection in human daily activity recognition », *in*: *2025 21th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT)*, IEEE, Tuscany, Italy, June 2025.

[221] Davide Anguita et al., « A public domain dataset for human activity recognition using smartphones. », *in*: *Esann*, vol. 3, 1, 2013, pp. 3–4.

[222] George Vavoulas et al., « The mobifall dataset: Fall detection and classification with a smartphone », *in*: *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)* 2.*1* (2014), pp. 44–56.

[223] Angela Sucerquia, José David López, and Jesús Francisco Vargas-Bonilla, « SisFall: A fall and movement dataset », *in*: *Sensors* 17.*1* (2017), p. 198.

[224] Haruyuki Ichino et al., « HASC-PAC2016: Large scale human pedestrian activity corpus and its baseline recognition », *in*: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 705–714.

[225] Alexander Jamieson et al., « Human activity recognition of individuals with lower limb amputation in free-living conditions: A pilot study », *in*: *Sensors* 21.*24* (2021), p. 8377.

[226] Mansour Tchenegnon et al., « Handi-Motion: Corpus de Données de Mouvements Capturés en Situation de Handicap », *in*: *Handicap 2024*, 2024.

[227] Li Wang et al., « A privacy-preserving framework with multi-modal data for cross-domain recommendation », *in*: *Knowledge-Based Systems* 304 (2024), p. 112529.

[228] Akey Sungheetha et al., « MM-HealthGuard: Multimodal Deep Learning Framework

\

eta_ {\text {health}} $forReal-TimeHealthMonitoringwithPrivacy-PreservedStreamPr$
in: *2025 3rd International Conference on Inventive Computing and Informatics (ICICI)*, IEEE, 2025, pp. 789–794.

Titre : Assurer la protection de la vie privée lors de l'externalisation des données par des applications mobiles et IoT

Mot clés : protection de la vie privée, les objets connectés, sélection de caractéristiques préservant la confidentialité, données temporelles, personnes en situation de handicap, jeux de données HAR inclusifs

Résumé : Ces travaux portent sur la protection de la vie privée face au développement des technologies IoT et de l'IA dont l'usage croissant dans les environnements d'assistance pour les personnes en situation de handicap entraîne de nouveaux risques liés aux données personnelles. Les informations issues des capteurs peuvent révéler des éléments sensibles sur l'identité ou la santé d'une personne. Menée dans le cadre du projet France 2030 HIT, cette thèse conçoit des mécanismes protégeant les données lors de leur externalisation et de leur exploitation. Les travaux se concentrent sur les données temporelles utilisées pour la reconnaissance d'activités et proposent une sélection de caractéristiques basée sur un algorithme génétique à objectifs multiples. Les caractéristiques utiles sont ainsi retenues tout en écartant celles susceptibles d'exposer l'identité d'une personne. Le traitement est réalisé en edge computing afin de limiter l'exposition des données. L'approche maintient la performance des modèles d'apprentissage tout en réduisant les risques de ré-identification, y compris en cas d'accès non autorisé aux données annotées. Les travaux ont été étendus pour résister à des attaquants plus puissants. Les expériences s'appuient sur des bases de reconnaissance d'activités humaines, les données relatives aux personnes handicapées n'étant pas disponibles actuellement. Dans ce contexte, un jeu de données a été collecté avec le Centre Mutualiste de Kerpape dans un appartement domotisé ; des résultats préliminaires montrent qu'il est possible de distinguer les participants handicapés et valides à partir des signaux collectés, ce qui confirme le risque pour la vie privée.

Title: Ensuring privacy protection in data outsourcing by mobile applications and IoT

Keywords: privacy protection, Internet of Things, privacy-preserving feature selection, time series data, people with disabilities, inclusive HAR datasets

Abstract: This work focuses on the protection of privacy in the context of the development of IoT and AI technologies, whose growing use in assistive environments for people with disabilities introduces new risks related to personal data. Information collected from sensors can reveal sensitive elements concerning a person's identity or health. Conducted within the framework of the France 2030 HIT project, this thesis designs mechanisms to protect data during their outsourcing and processing. The work focuses on temporal data used for activity recognition and proposes a feature selection approach based on a multi-objective ge-

netic algorithm. Useful features are retained, while those that could reveal a person's identity are discarded. Processing is performed through edge computing in order to limit data exposure. The approach maintains the performance of learning models while reducing the risk of re-identification, including in cases of unauthorized access to annotated data. The work has been extended to withstand more powerful attackers. Experiments rely on human activity recognition datasets, as data concerning people with disabilities are not currently available. In this context, a dataset was collected in collaboration with the Centre Mutualiste de Kerpape in a smart apartment; preliminary results show that it is possible to distinguish between disabled and non-disabled participants from the collected signals, highlighting the associated privacy risk.