Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.066270

ARTICLE

Adversarial Perturbation for Sensor Data Anonymization: Balancing Privacy and Utility

Tatsuhito Hasegawa*,# and Kyosuke Fujino#

Graduate School of Engineering, University of Fukui, Fukui, 910-8507, Japan

*Corresponding Author: Tatsuhito Hasegawa. Email: t-hase@u-fukui.ac.jp

#These authors contributed equally to this work

Received: 03 April 2025; Accepted: 29 May 2025; Published: 03 July 2025

ABSTRACT: Recent advances in wearable devices have enabled large-scale collection of sensor data across healthcare, sports, and other domains but this has also raised critical privacy concerns, especially under tightening regulations such as the General Data Protection Regulation (GDPR), which explicitly restrict the processing of data that can re-identify individuals. Although existing anonymization approaches such as the Anonymizing AutoEncoder (AAE) can reduce the risk of re-identification, they often introduce substantial waveform distortions and fail to preserve information beyond a single classification task (e.g., human activity recognition). This study proposes a novel sensor data anonymization method based on Adversarial Perturbations (AP) to address these limitations. By generating minimal yet targeted noise, the proposed method significantly degrades the accuracy of identity classification while retaining essential features for multiple tasks such as activity, gender, or device-position recognition. Moreover, to enhance robustness against frequency-domain analysis, additional models trained on transformed (e.g., short-time Fourier transform (STFT)) representations are incorporated into the perturbation process. A multi-task formulation is introduced that selectively suppresses person-identifying features while reinforcing those relevant to other desired tasks without retraining large autoencoder-based architectures. The proposed framework is, to our knowledge, the first AP-based anonymization technique that (i) defends simultaneously against time- and frequency-domain attacks and (ii) allows per-task trade-off control on a single forward-back-propagation run, enabling real-time, on-device deployment on commodity hardware. On three public datasets, the proposed method reduces person-identification accuracy from 60–90% to near-chance levels (≤5%) while preserving the original activity-recognition F1 both in the time and frequency domains. Compared with the baseline AAE, the proposed method improves downstream task F1 and lowers waveform mean squared error, demonstrating a better privacy-utility trade-off without additional model retraining. These findings underscore the effectiveness and flexibility of AP in privacy-preserving sensor-data processing, offering a practical solution that safeguards user identity while retaining rich, application-critical information.

KEYWORDS: Human activity recognition; privacy-aware IoT; adversarial perturbation

1 Introduction

Wearable devices have rapidly evolved in recent years, enabling continuous and large-scale collection of sensor data related to human activity, physiology, and environment. These sensors, embedded in smartwatches, smartphones, or fitness bands, capture diverse signals such as acceleration, heart rate, and gyroscopic measurements. Such capabilities have facilitated breakthroughs in healthcare applications [1–3], sports science [4], and human-computer interaction [5]. However, the increased availability and granularity of sensor data have also raised critical privacy concerns [6,7]. Sensitive attributes, ranging from demographic

information to health indicators, can be inferred from seemingly harmless motion signals when advanced machine learning techniques are applied [8,9].

To mitigate the risk of re-identification and protect sensitive personal attributes, various anonymization frameworks have been proposed [10,11]. One well-studied approach is the Anonymizing AutoEncoder (AAE) [12,13], which integrates autoencoder-based transformations with a supervised loss term to degrade identity-related features while preserving a target classification task (e.g., Human Activity Recognition; HAR). While AAE can effectively lower identification accuracy, it often introduces substantial distortions in the waveform, reducing the data's utility for tasks beyond the one explicitly used in training the autoencoder. Moreover, it assumes that new analysis models will be trained on the anonymized data itself, making it difficult to reuse established models trained on non-anonymized (raw) signals.

This study proposes a new sensor data anonymization framework Anonymizing Adversarial Perturbation (AAP), which applies subtle perturbations to inputs, leveraging adversarial perturbations (AP) [14], to balance privacy and multi-task utility. The proposed method uses identity classification models to generate targeted, minimal distortions that degrade re-identification accuracy, while concurrently reinforcing or preserving important features for other tasks such as gender or detailed HAR. This extended scheme is referred to as Frequency-informed AAP (F-AAP) and Multi-task Frequency-informed AAP (MF-AAP). Through extensive experiments on publicly available sensor datasets (Motion Sense [15], MHEALTH [16], and UniMiB SHAR [17]), this study demonstrates that:

1. The proposed adversarial-perturbation approach preserves richer waveform characteristics than autoencoder-based anonymization, enabling higher accuracy on tasks not explicitly considered during anonymization.
2. Incorporating multiple models trained in time and frequency domains leads to greater resilience against re-identification, even if adversaries transform the signals using short-time Fourier Transform (STFT) methods.
3. A multi-task formulation allows users to selectively strengthen or weaken different task-related features with no need to retrain large generative networks, greatly improving flexibility in real-world deployments.

Although anonymization methods based on autoencoders [12] and Generative Adversarial Networks (GANs) [11] can obscure user identity, they require large models and retraining and often degrade downstream–task accuracy. Moreover, prior work implicitly assumes that an attacker operates in the time domain, overlooking the fact that simple spectral transforms can re-expose user-specific cues. To date, no study has applied adversarial perturbations to sensor signals while simultaneously preserving the utility of multiple downstream tasks. This gap motivates the present work, which introduces AAP, F-AAP, and MF-AAP to (i) anonymize wearable-sensor data with minimal waveform distortion, (ii) remain robust in both time and frequency domains, and (iii) retain high accuracy for activity, position, and gender recognition. Overall, the adversarial-perturbation methodology offers a lightweight yet powerful alternative to AAE-based sensor data anonymization. By focusing on minimal and targeted modifications, it achieves strong privacy guarantees while maintaining high utility across diverse tasks, which is essential for the next generation of wearable sensing systems.

The objectives of this study are as follows:

O1 Reduce person-identification accuracy to chance level while introducing minimal waveform distortion.
O2 Achieve robustness against spectral attacks by leveraging both time- and frequency-domain models.
O3 Enable selective utility preservation for multiple downstream tasks without retraining large networks.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 defines the privacy-preserving scenario of this study; Section 4 details the proposed AAP, F-AAP, and MF-AAP; Section 5 presents experimental setup and results; Section 6 discusses limitations and future work; finally, Section 7 concludes the paper.

2 Related Work

Sensor data have been extensively leveraged in diverse domains, ranging from industrial settings and urban infrastructures to healthcare and daily life. Nevertheless, such broad applicability has also prompted critical discussions on privacy and security. This section first introduces representative use cases of sensor data, followed by an overview of HAR methodologies. Existing anonymization techniques, including conventional statistical methods, GAN-based approaches, and autoencoder-based methods, are surveyed, after which recent advances in adversarial perturbation for privacy are discussed. The present study is then positioned in relation to these prior works.

2.1 Applications of Sensor Data

Sensor deployments in industrial IoT environments have enabled real-time monitoring and optimization of manufacturing processes. Xu et al. [18] proposed a hierarchical resource allocation algorithm that takes into account safety, privacy, and reliability constraints, thereby enhancing efficiency across industrial operations. In the context of smart cities, Talebkhah et al. [19] reported ongoing projects that utilize sensor networks for traffic control, environmental surveillance, and disaster management. These initiatives highlight the transformative potential of large-scale sensor deployments in urban planning and sustainability.

Wearable technologies are also finding widespread adoption in sports science and consumer fitness. Lam Po Tang [4] demonstrated how wearable sensors collecting heart rate and posture data can be harnessed to refine training regimens and speed recovery. Meanwhile, newly developed smart garments and textiles are emerging to measure physiological signals in everyday settings. In healthcare, sensor-based HAR has proven beneficial for patient care and clinical efficiency. Lee et al. [20] introduced a lifelogging system employing three-axis accelerometers and combined statistical and spectral features to achieve high-accuracy HAR for daily-life analysis. Similarly, Xu et al. [21] exploited random forest classifiers augmented with contextual information to recognize activities among seniors. Inoue et al. [22] developed a system for analyzing nursing workflows, aiming to improve operational efficiencies in clinical environments. Sensor technologies have likewise progressed across a broad spectrum, encompassing systems for pain monitoring and mitigation [3] as well as implantable *in-vivo* sensors [2]. While these studies underscore the utility of sensor data, they also underline the growing importance of privacy protection to secure personal information against misuse.

2.2 HAR

HAR encompasses a range of techniques aimed at classifying sensor signals into specific behavioral categories. Traditional HAR approaches predominantly employed handcrafted statistical or spectral features, which were then input into machine learning algorithms. Kwapisz et al. [23] used smartphone accelerometers to identify walking and stair-climbing behaviors, demonstrating how fundamental statistical features can boost model accuracy. By contrast, Shoaib et al. [24] combined data from smartphones and wristbands to handle more complex activities and highlighted the value of sensor diversity. Voicu et al. [25] integrated readings from accelerometers, gyroscopes, and gravity sensors, showcasing the feasibility of accurate activity classification solely using commercial smartphones. There are also cases where accelerometers have been applied to enhance the security of voice authentication [26].

In more recent developments, deep learning models, particularly convolutional neural networks and recurrent neural networks, have been adopted in an end-to-end manner. Li et al. [27] found that deep neural networks outperform methods reliant on hand-crafted features when analyzing wearable sensor data. DenseNet-inspired architectures have also been proposed to capture spatiotemporal dependencies effectively, as exemplified by Ronald et al. [28] in their HARDenseNet. Ronald et al. [29] integrated ResNet and Inception modules to develop iSPLInception, achieving high accuracy in HAR tasks. In recent years, efforts to leverage Transformer architectures for human activity recognition have also become widespread [30–32]. While deep learning has propelled the performance of HAR systems, it has also amplified privacy concerns, because large-scale sensor data collection can reveal sensitive personal details.

2.3 Sensor Data Anonymization

In the field of sensor data anonymization for HAR, several comprehensive surveys have been conducted [33,34]. Based on these surveys, existing algorithm-based anonymization methods can be broadly classified into three categories: (a) Statistical Anonymization, (b) Generation-based Anonymization, and (c) Reconstruction-based Anonymization.

2.3.1 (a) Statistical Anonymization Techniques

Several statistical approaches exist for sensor data anonymization, each attempting to mask identity-revealing details while preserving data utility. Filtering eliminates identifiable patterns in sensor signals through time- or frequency-domain transformations [35]. For instance, removing certain frequency bands may obscure users' unique motion signatures. However, this strategy risks losing salient information necessary for downstream tasks. Data perturbation injects random noise to conceal personal features without fully distorting the dataset's broader statistical properties. Gaussian noise addition [36] is a commonly cited example. Striking a balance is vital: excessive noise diminishes data utility, while insufficient noise leaves identifying cues intact.

Data generalization replaces precise readings with coarser-grained versions. For example, converting timestamps to approximate time bins or rounding sensor measurements to broader intervals can mitigate re-identification risks [37]. By introducing k-anonymity [38], each data record becomes indistinguishable from at least $(k-1)$ others, but at the cost of reduced specificity for analytics [39]. Differential privacy provides a formal mechanism to limit the impact of any single record on aggregate statistics. However, implementing it in sensor-based HAR remains nontrivial, since ensuring robust privacy often requires a high level of noise, thereby degrading recognition accuracy [10]. Random Projection (RP) [40] is a technique that projects high-dimensional data into a randomly chosen lower-dimensional subspace, leveraging the Johnson-Lindenstrauss lemma to approximately preserve pairwise distances. This property hampers direct reidentification attempts because reconstructing the original sensor signals becomes more difficult. However, when the projection dimension is chosen with care, the sensor data can still retain sufficient utility for subsequent recognition tasks such as classification or clustering. A trade-off nevertheless remains: an overly aggressive reduction in dimensionality may obscure features essential for analytics, whereas a projection that is too large may continue to expose identifying signatures.

While these statistical methods are relatively straightforward, they each face inherent constraints, especially when aiming to maintain the fidelity necessary for sophisticated HAR tasks.

2.3.2 (b) Generation-Based Anonymization

GANs [41] have emerged as a promising tool for synthesizing sensor data that preserve certain target attributes while concealing sensitive ones. Menasria et al. [11] introduced Private GAN (PGAN) frameworks (PGAN1 and PGAN2), focusing on selectively safeguarding private attributes and retaining public information. By generating data from a learned distribution rather than sharing direct measurements, the approach lessens re-identification risks. In addition, anonymization methods based on the GAN with conditional AE [42] and approaches that combine GANs with microaggregation [43] have also been proposed. However, GAN-based methods often demand large-scale datasets and can be difficult to tailor for individualized user privacy.

Some studies have tackled anonymization by applying adversarial training (AT) in the feature space without generating waveforms [44,45]. These methods can be regarded as approaches for generating an anonymized feature space. Furthermore, anonymization methods based on diffusion models have also been proposed in recent years [46].

2.3.3 (c) Reconstruction-Based Anonymization

Malekzadeh et al. [12] proposed an Anonymizing AutoEncoder (AAE) that simultaneously degrades user identification accuracy and retains utility for a designated recognition task (e.g., HAR), and Bigelli et al. [13] extend AAE for preventing some privacy attributes. AAE functions by transforming each sensor sample using an autoencoder constrained by classification losses for user ID and activity, alongside a mean squared error (MSE) term to mitigate distortion. Although AAE generally performs well for its intended use case (i.e., training new HAR models on anonymized data), it can exhibit considerable waveform modifications (Fig. 1) and lacks explicit mechanisms for preserving information relevant to tasks other than the primary classification. Moreover, anonymity under frequency-domain analysis remains insufficiently addressed.

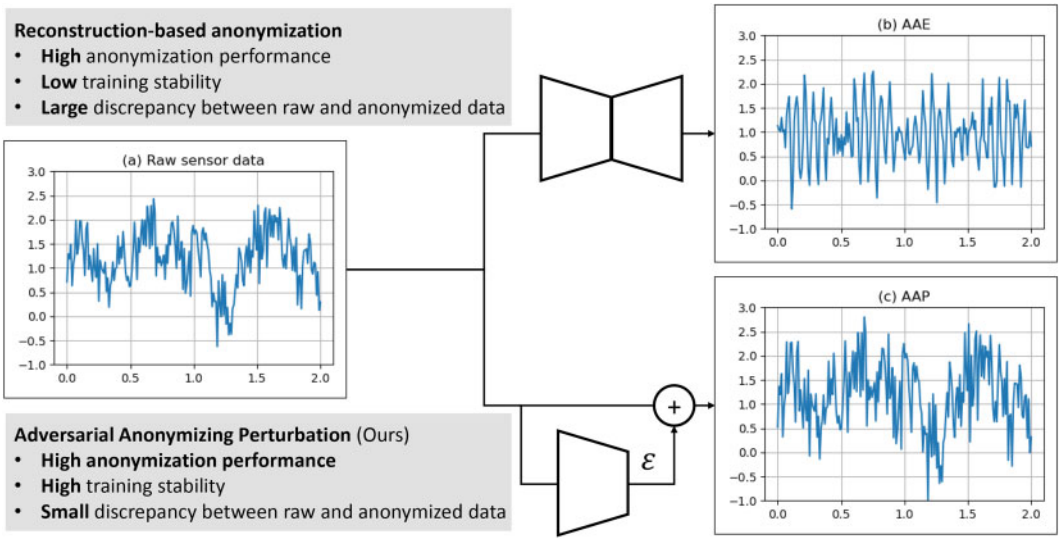

Figure 1: Samples of waveform changes before and after conversion using the reconstruction-based method and the proposed method

As a related approach, in image-based human activity recognition, autoencoder-based method [47] have been proposed to address reconstruction-based threats. An integrative method utilizing autoencoders has also been proposed [48].

2.3.4 AP for Privacy

AP [14] were initially studied in computer vision and speech recognition as imperceptible noise that induces misclassifications. Strategies such as Basic Iterative Method (BIM) [49], Diverse-Inputs Iterative Fast Gradient Sign Method (DI^2-FGSM) [50], and Translation-Invariant FGSM (TI-FGSM) [51] refine the basic FGSM [52] approach to enhance attack strength or transferability by iterating over gradient updates or employing transformations and smoothing filters [53]. Outside of pure security contexts, these methods can be repurposed for anonymization: minor signal distortions strategically undermine identity classification while leaving much of the original data structure intact.

Although AP have been validated in image and audio domains, their application to time-series sensor data, particularly for privacy protection, remains an emerging area. AP-based anonymization can offer advantages over generative or statistical approaches by requiring minimal structural adjustments to raw signals.

2.4 Positioning of the Present Work

This section clarifies the positioning of the proposed methods relative to existing anonymization approaches. Table 1 compares six representative methods using seven columns, each reflecting a key property of sensor-data anonymization techniques.

Table 1: Characteristics comparison of various anonymization methods

Method	Ann.	Stab.	Anon.	Pres.	Int. adj.	Freq.	Flex.	HAR
Statistical	None	High	Middle	Middle	✓	✗	✗	✗
GAN-based [11]	Personal	Low	High	Low	✗	✗	✗	✗
AT-based [44,45]	Personal	Middle	High	Low	✗	✗	✓	✓
AAE [12,13]	Personal	Middle	High	Low	✗	✗	✗	✓
AAP	Personal	High	High	High	✓	✗	✗	✓
F-AAP	Personal	High	High	High	✓	✓	✗	✓
MF-AAP	Various info.	Middle	High	High	✓	✓	✓	✓

2.4.1 Abbreviations and Their Meanings

- **Method:** The name or category of each anonymization approach.
- **Ann. (Requiring annotation):** Indicates whether annotation labels are required to train the anonymization model. While GAN-based and autoencoder-based methods use personal labels to *maximize* the personal identification loss (for training a discriminator or autoencoder), AAP uses them to *minimize* the personal identification loss. In addition, both AAE and AAP rely on target labels (e.g., activity labels) to enhance human activity recognition (HAR) performance.
- **Stab. (Stability):** The extent to which an approach preserves the original waveform structure. *High* means minimal distortion, *Middle* indicates moderate change, and *Low* suggests a high degree of alteration.

- **Anon. (Anonymity):** The effectiveness in degrading person-identification accuracy (i.e., increasing re-identification difficulty). *High* implies strong anonymization (significantly reducing user-specific signals), *Middle* indicates partial anonymization, and *Low* suggests limited success in concealing identity.
- **Pres. (Information preservation):** How effectively each method retains task-relevant information in the anonymized data. *High* indicates minimal loss of essential features, *Middle* indicates moderate loss, and *Low* indicates that critical signals needed for downstream tasks may be severely disrupted.
- **Int. adj. (Intensity adjustability):** Whether the method offers flexible control over anonymization strength. GAN-based and autoencoder-based methods typically make re-identification more difficult *only* during the training phase; thus, their anonymization strength is not easily adjustable at inference time. By contrast, Statistical and AAP approaches can tune noise intensity to balance privacy and utility on demand.
- **Freq. (Frequency-domain robustness):** Whether the method remains effective against frequency-domain analysis. A check mark (✓) signifies that the anonymization withstands transformations such as STFT, whereas a blank cell indicates vulnerability to frequency-based attacks.
- **Flex. (Flexibility):** The ability to selectively preserve or anonymize different attributes of the data. MF-AAP, for example, can flexibly determine which characteristics (e.g., gender or device position) are retained and which are suppressed.
- **HAR:** Whether or not consideration is given to maintaining behavior recognition accuracy.

2.4.2 Comparative Features of the Proposed Approach

As shown in Table 1, traditional Statistical methods generally do not require unique annotation labels but do not explicitly anonymize user attributes. Although they allow parameter tuning (Int. adj. ✓), their anonymization and information preservation capabilities remain at only moderate levels (Anon. = Middle, Pres. = Middle). By contrast, GAN- and AT-based techniques can strongly reduce user identification risk (Anon. = High), as with differential privacy. However, no prior work has been found that leverages GANs to maintain or improve HAR accuracy. Moreover, they demand large training datasets and specialized hyperparameter tuning for stable operation (Stab. = Low) and tend to lose fine-grained information for multiple tasks (Pres. = Low).

AAE (Anonymizing AutoEncoder) reaches high anonymity (Anon. = High) while preserving certain targeted behaviors, but it requires strict hyperparameter tuning for stabilising model training because of using the minimax optimization such as GANs (Stab. = Middle) and has limited capacity for retaining diverse information (Pres. = Low). Additionally, most autoencoder-based approaches do not provide an explicit intensity parameter for fine-tuning (Int. adj. is blank), nor do they address frequency-based vulnerabilities.

The proposed methods address these limitations in a stepwise manner. AAP introduces AP specifically targeting user identity signals while preserving the waveform structure (Stab. = High) and essential features for various tasks (Pres. = High). Under the new column definition, AAP requires both personal labels and specific target labels for training (Ann. = Personal & Target), thereby allowing it to degrade identification accuracy while keeping task-relevant information (e.g., activity). Unlike AAE, it retains a check mark for Int. adj. by allowing users to tune parameters such as intensity of perturbations. However, basic AAP does not include explicit defenses in the frequency domain (Freq. column is blank).

F-AAP extends AAP by integrating frequency-domain models (Freq. ✓), enabling it to maintain high anonymity (Anon. = High) even under STFT-based analysis, while still preserving waveform stability (Stab. = High) and crucial task information (Pres. = High). In terms of annotation, F-AAP remains similar to AAP (Ann. = Personal & Target), but now provides robust anonymization in both time and frequency domains.

Finally, MF-AAP broadens its training requirements to cover various user attributes alongside the main classification targets (Ann. = Various info. & Target). This multi-task design preserves complex task-specific features (Pres. = High) and achieves strong anonymity (Anon. = High), though the added complexity can slightly reduce stability of model training a little (Stab. = Middle). Moreover, MF-AAP carries a check mark in every remaining category, including Flex. (✓) for selectively suppressing certain user attributes (e.g., gender) while retaining others (e.g., activity). This flexibility allows practitioners to adaptively tune or reinforce features relevant to different downstream applications.

In summary, AAP, F-AAP, and MF-AAP collectively surpass prior methods in balancing anonymity and utility, offering explicit adjustability, multi-task preservation, and robust frequency-domain defenses. These advantages position the proposed methods as promising solutions for high-fidelity sensor-data anonymization across a wide range of use cases. Building on the above comparison, the distinct novelties of this study are: (i) a new adversarial-perturbation paradigm that dispenses with GAN/auto-encoder architectures, (ii) the first sensor-signal anonymization defence that is simultaneously robust in both time and frequency domains (F-AAP), (iii) an attribute-selective privacy–utility controller that preserves multiple downstream tasks without retraining (MF-AAP), and (iv) comprehensive cross-dataset evidence demonstrating a new state-of-the-art privacy–utility trade-off.

3 Privacy-Preserving Scenario

3.1 Scenario and Elements

This section presents the overall privacy-preserving scenario assumed in this study, along with the requirements that must be satisfied in this context. As illustrated in Fig. 2, the system adopts a server–client architecture for collecting and utilizing sensor data obtained from smartphones and wearable devices. The scenario consists of the following four elements:

User devices

Users employ smartphones or wearable devices to measure the sensor data, attaching the corresponding activity labels before transmitting the data to the application server. Rather than sending the raw sensor data directly, the user devices perform anonymization locally. In doing so, any information enabling personal identification is blocked at the source, preventing raw data from ever reaching the server.

Application server

The application server aggregates and manages the anonymized data and activity labels from multiple user devices, offering services such as lifelogging or other sensor-driven applications. In addition, the server may provide the collected anonymized data to approved third parties for further utilization—e.g., activity analysis or the training of activity-recognition models. However, since no personally identifying information is transmitted, the risk of linking data to specific individuals on the server side is minimized.

Data consumers

After receiving anonymized data from the server, data consumers perform various tasks such as activity classification or in-depth behavioral analysis. Additional use cases are anticipated through transfer learning or self-supervised approaches, wherein the anonymized data may be repurposed for tasks beyond basic activity recognition.

Attacker

It is assumed that potential adversaries may attempt to illegally acquire data from the server via methods such as malware, man-in-the-middle (MitM) attacks, or phishing. Particularly problematic is the case where attackers hold a pre-trained person-identification model based on previously leaked raw

data. Even if the data on the server are anonymized, the attacker could re-identify individuals if the anonymization is insufficient. This research therefore aims to degrade identification accuracy significantly through on-device anonymization.

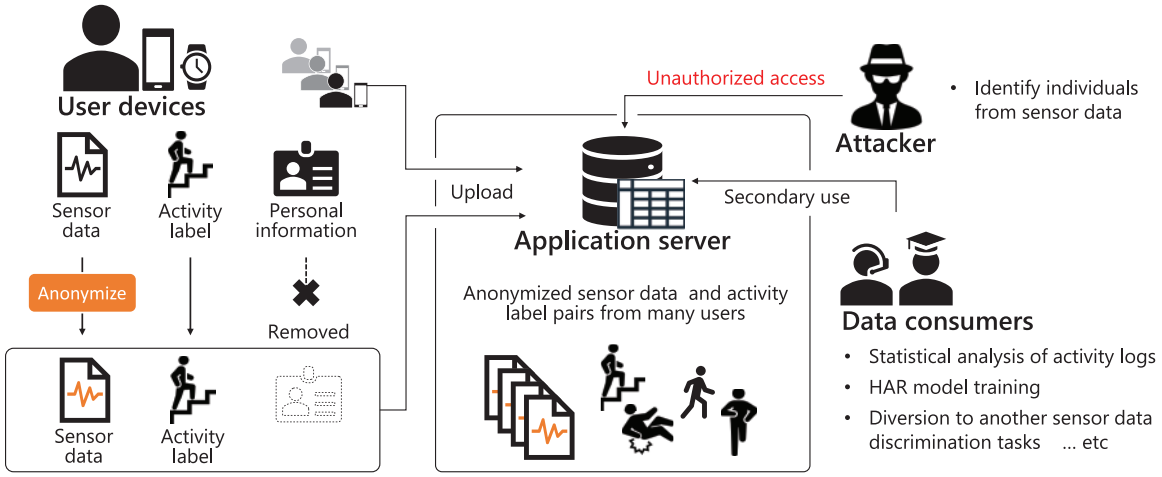

Figure 2: Overview of the assumed privacy-preserving scenario

3.2 Risk Examples Based on the Attacker's Possessed Information

Under the conditions of this scenario, attackers may possess the following types of information, which can lead to different privacy risks:

(a) Application server login credentials

By impersonating legitimate users or administrators, attackers can access the anonymized data and associated activity labels on the server. However, since direct personal identifiers are absent, re-identification risk remains low unless the attacker can cross-reference external sources of raw sensor data.

(b) Personal identification model + (a)

In addition to (a), by using a person-identification model trained on previously leaked raw sensor data, there is a risk that an attacker can link the sensor data in the server to specific individuals if the anonymization is insufficient. This enables them to correlate behaviors with identified users.

(c) Raw sensor data + personal ID + (a)

If the attacker already has direct access to users' sensor data and personal IDs from some other breach, they can build or refine a personal identification model and pose essentially the same threat as in case (b). In addition, if the anonymization method has also leaked, they can reproduce the anonymized sensor data and build the user identification model supporting anonymized sensor data.

3.3 Anonymization Requirements

Generally, anonymization in data handling is expected to address the following five points:

(1) Removal or masking of personally identifying information
(2) Reduction of re-identification risk
(3) Preservation of data utility
(4) Compliance with relevant laws and regulations
(5) Transparency and accountability

In the scenario of this study, requirement (1) is already addressed on the user device side by design, and (4) and (5) fall under policy or operational guidelines. Hence, for sensor data anonymization, the primary concerns are (2) reducing re-identification risk and (3) preserving data utility.

Past research has proposed many anonymization methods that obscure personally identifying information while retaining features vital for tasks such as activity classification [12]. However, when the transformed data deviate substantially from the original waveform, important information for secondary use cases may be lost. For instance, if only the "activity label" is retained, the raw waveform's additional characteristics, potentially valuable for other analyses, become inaccessible. Therefore, the objective is to degrade certain specific aspects of the data (namely person-identification signals) while still preserving as much of the original data characteristics as possible. On the other hand, it should be noted that these are trade-offs.

Based on the above considerations, three requirements emerge for anonymization in the scenario of this study:

- Transform the data such that a person-identification model trained on real (raw) data can only achieve chance-level accuracy when applied to the anonymized data.
- Retain information necessary for key tasks, particularly activity recognition, so that classification performance remains sufficiently high.
- Preserve diverse features to support broader use, such as other classification tasks (e.g., position estimation) beyond simple activity analysis.

By satisfying these requirements, even attackers armed with person-identification models built from leaked raw data will be significantly hampered in re-identifying individuals. Meanwhile, data consumers can still utilize the anonymized data for activity recognition and new downstream applications, thus helping to reduce user reluctance to share sensor data in an era of heightened privacy awareness.

4 Proposed Method

This study proposes a new adversarial-perturbation-based approach to supplement the limitations of existing anonymization methods such as AAE, which often suffer from excessive waveform distortions or focus on only a specific task. For instance, in the context of person images, AP can reduce classification accuracy of a face-recognition model while preserving the visual appearance. However, because these visual changes are quite subtle, a person-recognition model might be deceived, yet humans can still identify the individual by simple inspection, hence it fails to achieve anonymization. In contrast, when data are inherently difficult to identify visually, as with sensor signals, the act of degrading the classification model's accuracy itself effectively serves as anonymization. Thus, in the field of activity recognition, where human observation cannot easily detect identities, AAP leverages this unique property of sensor data. No prior reports have been identified that employ AP to anonymize sensor data, suggesting that the present approach opens a new direction for anonymization research.

This section first introduces AAP, which applies AP in the time domain to reduce person-identification accuracy while minimizing waveform distortion, thereby explaining the fundamental process of the proposed framework. The approach is then extended to F-AAP, which combines time-domain and frequency-domain models to ensure that anonymization remains robust in the frequency domain. Finally, it is further extended to MF-AAP, which simultaneously considers multiple tasks (e.g., device-position estimation) and selectively degrades only person-identification accuracy. Across all proposed methods, the input and output formats follow those of the related study [12]. Specifically, sensor waveforms segmented into fixed-length windows by a sliding-window preprocessing step are supplied as input, and anonymized

waveforms of the same length are produced as output. No additional preprocessing is applied. The concrete experimental settings are described in Section 5.

4.1 Anonymization with a Simple Adversarial Perturbation (AAP)

An anonymization scheme must balance privacy protection (i.e., lowering person identification accuracy) with data utility (i.e., retaining essential information). Although increasing waveform distortion can enhance anonymization, it risks destroying the inherent features of the data. AP, on the other hand, can significantly disrupt a classifier's inference while introducing only a minimal visible change to the data. Inspired by AP methods developed in the image domain, the technique is adapted to sensor data in order to preserve the original waveforms as much as possible while drastically reducing person-identification accuracy, and this variant is referred to as AAP.

Fig. 3 illustrates the flow of AAP. The method employs IFGSM [49], which iteratively applies the AP by FGSM [52]. In advance, a person-identification model (M_{id}) is trained on the raw data (with parameters θ_{id}). At anonymization time, the loss gradient $\nabla_{\boldsymbol{X}}$ with respect to the input sensor data $\boldsymbol{X}$ and the person label $\boldsymbol{y}_{\mathrm{id}}$ is computed and incrementally added t times to produce the anonymized data $\boldsymbol{X}_t$. Eq. (1) shows an example transformation step at iteration t.

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} + \alpha \,\mathrm{sign}\Big(\nabla_{\boldsymbol{X}} J_{\theta_{\mathrm{id}}}(\boldsymbol{X}_{t-1}, y_{\mathrm{id}})\Big). \tag{1}$$

Here, $\boldsymbol{X}_t$ is the anonymized data at the t-th iteration, and $J_{\theta_{\mathrm{id}}}(\cdot)$ is the loss function of the person-identification model. The final magnitude of waveform change depends on parameters α and the number of iterations t. In this context, α denotes the step size added at each iteration. Unlike prior adversarial-attack studies, this work does not impose an explicit L_p budget. Instead, the step size α is selected via grid search so that the person identification F1 scores drop to chance level. Unlike autoencoder-based methods (e.g., AAE), AAP does not reconstruct the original waveform, only minor perturbations are added. This property allows the scheme to degrade classification accuracy using small noise while offering a means to fine-tune the degree of anonymization at deployment by adjusting these parameters.

While AAE encodes waveforms into latent variables and then reconstructs them, often yielding a significant gap between original and reconstructed data, AAP only adds minimal noise to reduce identification accuracy. Consequently, one may expect that additional information (e.g., frequency characteristics) remains more readily preserved under AAP. Subsequent experimental evaluations (see Section 5) demonstrate that AAP offers broader data utility than AAE.

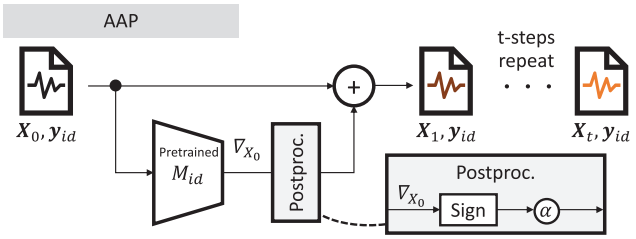

Figure 3: Anonymization process by AAP. AAP takes the sensor waveform $\boldsymbol{X}$ and the person label $\boldsymbol{y}_{\mathrm{id}}$ as input, computes the gradient $\nabla_{\boldsymbol{X}}$ with respect to $\boldsymbol{X}$ using a pretrained person-identification model M_{id}, and adds this gradient to the original waveform as a perturbation

4.2 *Frequency-Informed Adversarial Perturbation for Anonymization (F-AAP)*

AAP focuses on classification models in the time domain. However, if an attacker trains a person-identification model in the frequency domain, standard AAP might fail to anonymize effectively. Since sensor data (e.g., accelerometry or biosignals) often contain unique frequency components tied to individual users, restricting anonymization efforts solely to the time domain can be insufficient.

F-AAP is introduced, which concurrently derives gradients from both time-domain and frequency-domain models to guide perturbations so that person-identification becomes difficult in both domains. As depicted in the left of Fig. 4, the sensor data $\boldsymbol{X}^{(\mathrm{T})}$ are transformed into the frequency domain $\boldsymbol{X}^{(\mathrm{F})}$ via STFT, and then combine the time-domain and frequency-domain losses. Eq. (2) shows that the gradients $\nabla_{\boldsymbol{X}^{(\mathrm{T})}} J_{\theta_{(\mathrm{T,id})}}$ and $\nabla_{\boldsymbol{X}^{(\mathrm{F})}} J_{\theta_{(\mathrm{F,id})}}$ are used, subsequently mapping back via ISTFT and aggregating the signs.

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} + \alpha \sum_{d \in \{\mathrm{T},\mathrm{F}\}} \delta_d \cdot \mathrm{sign}\left(\nabla_{\boldsymbol{X}} J_{\theta_{\mathrm{id}}^{(d)}}\left(\boldsymbol{X}_{t-1}^{(d)}, y_{\mathrm{id}}\right)\right). \tag{2}$$

Here, $d = \mathrm{T}$ denotes the time domain, and $d = \mathrm{F}$ denotes the frequency domain; setting $\delta_{\mathrm{T}} = 1$, $\delta_{\mathrm{F}} = 1$ leverages both gradients, whereas setting one of them to 0 reverts to standard AAP or a single-domain approach. By integrating these two domains, F-AAP hinders attacker-driven identification whether in the time or frequency domain, while still preserving essential information in the data.

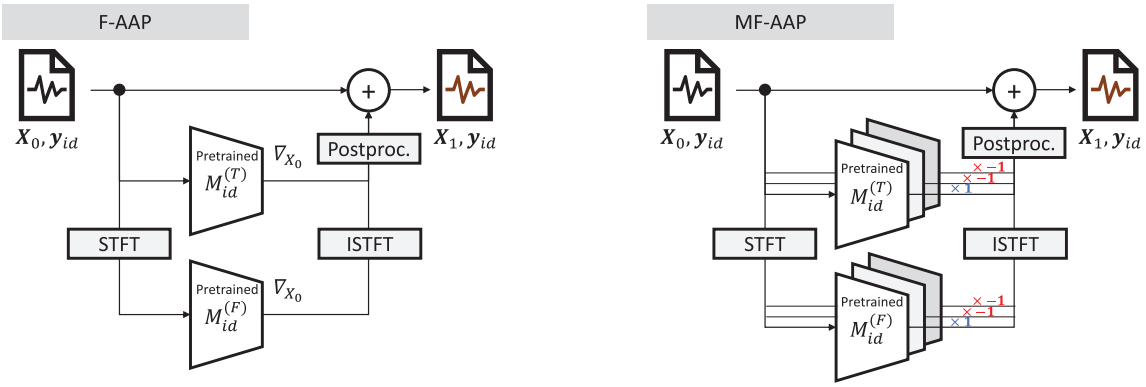

Figure 4: Anonymization processes by F-AAP (left) and MF-AAP (right)

4.3 *Adversarial Anonymization Considering Multiple Tasks (MF-AAP)*

Finally, beyond neutralizing person-identification models in the time and frequency domains, certain use cases may demand preservation of other tasks' accuracy (e.g., activity recognition, gender inference, or device-position estimation). For instance, a user may wish to obfuscate personal identity but keep activity recognition or gender prediction operational on the device.

To address such needs, F-AAP is extended to MF-AAP, which simultaneously handles "tasks that require accuracy preservation" (e.g., activity, gender, or position) and "tasks whose accuracy should be reduced" (i.e., person identification). As illustrated on the right of Fig. 4 and formalized in Eq. (3), each task is assigned a weight $\sigma_{\mathrm{task}} \in \{+1, 0, -1\}$, for example, $\sigma_{\mathrm{id}} = +1$ (intentionally lowering person-identification accuracy) and $\sigma_{\mathrm{act}} = -1$ (preserving or boosting the target task). Gradients are summed after applying these weights to produce a single perturbation that simultaneously degrades person identification while retaining accuracy for other tasks.

$$\boldsymbol{X}_t = \boldsymbol{X}_{t-1} + \alpha \sum_{d\in\{\mathrm{T,F}\}} \delta_d \cdot \mathrm{sign}\left(\sum_{\mathrm{task}\in\{\mathrm{id,act,...}\}} \sigma_{\mathrm{task}} \cdot \nabla_{\boldsymbol{X}} J_{\theta^{(d)}_{\mathrm{task}}}(\boldsymbol{X}^{(d)}_{t-1}, y_{\mathrm{task}}) \right). \tag{3}$$

Hence, MF-AAP can preserve the performance of tasks like activity classification while deliberately reducing only person-identification accuracy. Because one can set σ_{task} for each task independently, users can flexibly choose which information to protect and which to retain allowing them to tailor the anonymization to different deployment scenarios.

5 Evaluation Experiment

In this section, the three proposed methods (AAP, F-AAP, and MF-AAP) are experimentally evaluated against the following research questions (RQs) to address the gaps identified in the Introduction. RQ1 tests whether a perturbation (AAP) outperforms the prevailing AAE baseline in the privacy-utility trade-off; RQ2 examines whether adding a frequency-domain surrogate (F-AAP) preserves anonymity when an attacker operates in the spectral domain; and RQ3 validates that a multi-task extension (MF-AAP) can suppress only person-ID accuracy while preserving utility for other tasks such as activity and position recognition. Privacy is quantified by the macro-F1 score of person identification, whereas utility is measured by F1 (classification) or MSE (regression) on the downstream tasks.

- **RQ1:** Does the AP-based anonymization approach (AAP) outperform existing methods (e.g., AAE) in terms of both anonymization effectiveness and information preservation?
- **RQ2:** Is F-AAP effective at achieving robust anonymization in the frequency domain?
- **RQ3:** Can MF-AAP selectively suppress only person identification accuracy while preserving performance on other tasks?

5.1 Experimental Setup

5.1.1 Datasets and Pre-Processing

Three publicly available sensor datasets commonly employed in human activity and identity recognition are utilized:

- **Motion Sense** [15]: 24 subjects (gender-balanced), 6 activity classes.
- **MHEALTH** [16]: 10 subjects, 12 activity classes, sensors placed at 3 different body positions.
- **UniMiB SHAR** [17]: 30 subjects, 17 activity classes (9 daily-life and 8 fall-related activities).

All datasets contain sensor signals sampled at 50 Hz. Sliding windows (window size = 128, stride = 128) are applied to segment the data, followed by standardization. To assess anonymization effectiveness in both the time and frequency domains, the segmented data are transformed into spectrograms using STFT. Complex-valued spectrograms are represented by separately considering real and imaginary parts, effectively doubling the number of input channels.

In this paper, data from different domains are denoted by "T" for the time domain and "F" for the frequency domain. For example, the notations "Raw(T)", "Raw(F)", and "AAE(F)" are employed. Here, Raw(F) refers to data obtained by applying STFT to Raw(T). In contrast, AAE(F) does not denote STFT of AAE(T); rather, it represents data anonymized using an AAE model that has been trained in the frequency domain. The same definition applies to AAP(F).

5.1.2 Evaluation Tasks

Multiple classification tasks are defined to comprehensively evaluate anonymization and information retention capabilities:

- **Person Identification (Person ID):** Identifying subjects from sensor data.
- **Activity Recognition (Activity):** Classifying general activity types.
- **Sensor Position Estimation (Position):** Determining the sensor's location on the body.
- **Gender Recognition (Gender):** Classifying subjects' gender.
- **Detailed Activity Recognition (Detailed Act):** Classifying finely-grained activity categories.

The exact tasks and class numbers differ according to dataset characteristics.

5.1.3 Models and Training Procedure

A VGG-based architecture (VGG10) [54] is mainly employed as the primary evaluation model for all classification tasks, due to its established effectiveness and simplicity. To assess model transferability, additional experiments are conducted using a ResNet10 architecture [55], known for its robustness in deep learning literature.

The datasets are divided into training and test subsets (70% train, 30% test), ensuring balanced distributions of subjects and activities. The Adam optimizer with an initial learning rate of 0.001 is used for model optimization, and a cosine annealing scheduler progressively decreases the learning rate. The training runs for 500 epochs with a batch size of 128. Cross-entropy loss is used for training, with class-specific weighting to mitigate class imbalance effects. These hyperparameters were selected based on preliminary experiments.

The classification performance of AAP depends on both the constant α values and the number of perturbation iterations. In these experiments, the F1 score under settings where the person identification accuracy falls to or below the chance level are reported. To select this operating point, a grid search is performed over $\alpha \in \{0.0125, 0.025, \ldots, 0.25\}$, then choose the smallest α at which person-identification F1 scores dropped to or below the chance rate (e.g., 12.5% for eight classes), and finally evaluated all downstream tasks (activity, device-position, etc.) using that α, thus maximizing utility while satisfying the privacy constraint. The number of perturbation iterations is fixed at $t = 15$. The step size α is tuned in a sensitivity analysis and is held constant across all datasets in both the time and frequency domains.

5.1.4 Evaluation Metrics

Classification performance is evaluated using the mean F1 score averaged over five runs with different random seeds. The F1 score is a balanced metric based on the harmonic mean of precision and recall, well-suited to evaluate performance under class imbalance. To quantify changes introduced by anonymization, MSE between original and anonymized waveforms are calculated. Additionally, domain robustness (time/frequency) is evaluated using specialized metrics such as the id score and act score, detailed further in subsequent sections. These comprehensive evaluations allow us to compare the proposed approaches against the existing method (AAE) in terms of anonymization effectiveness and information preservation.

5.2 RQ1: Evaluation of the Effectiveness of AAP

5.2.1 Experimental Setup

In this section, the ability of AAP to simultaneously maintain high anonymity and improve information retention compared with the existing method (AAE) is evaluated. Anonymity is quantified by the degradation in person-identification performance, while information retention is measured by (i) the preservation

of classification accuracy on other tasks (e.g., activity, gender, and sensor-position recognition) and (ii) the magnitude of waveform change (quantified by the MSE between the original and anonymized signals). These metrics are compared comprehensively to assess the effectiveness of AAP.

The evaluation procedure is as follows:

1. For each dataset (Motion Sense, MHEALTH, and UniMiB SHAR), generate anonymized data by AAE and AAP.
2. Measure the classification performance (F1 score) of models built using VGG10 on these datasets (raw, AAE, and AAP).
3. Quantify the amount of waveform change by computing the MSE between the raw and anonymized data.

5.2.2 Evaluation via Classification Models

Table 2 shows the results. All models were trained using data in the time domain. The three leftmost columns denote the target estimation tasks, and the ten rightmost columns present the corresponding estimation results (mean F1 scores). The model architectures used are VGG10 and ResNet10, with VGG10 specifically employed for anonymization. "Raw," "RP [40]," "NOS2R2 [36]," "AAE [12]," "AAP," "F-AAP," and "MF-AAP" in the test data column indicate the test data and the anonymization methods applied.

Table 2: Comparison of anonymization performance on time-domain data [%]. In this scenario, the attacker has each model trained by time-domain sensor data

Model architecture			VGG10							ResNet10				
Test data			Raw	RP [40]	NOS2R2 [36]	AAE [12]	AAP	F-AAP	MF-AAP	Raw	AAE [12]	AAP	F-AAP	MF-AAP
Test data	Dataset	Class												
Person	Motion sense	24	56.7	0.6	3.3	1.8	2.4	3.7	3.4	58.5	1.6	3.3	4.1	4.0
	mHealth	10	85.1	2.5	7.7	4.3	2.5	2.5	4.8	86.8	5.3	8.8	9.3	9.2
	UniMiB	30	78.5	0.5	2.7	0.5	2.6	2.8	1.7	71.2	0.8	3.1	3.2	2.6
Activity	Motion Sense	6	80.2	6.8	40.9	40.6	79.0	79.7	99.6	81.8	37.7	63.6	64.7	74.6
	mHealth	12	82.8	1.6	20.0	10.9	71.8	70.0	90.5	80.4	9.7	59.4	61.0	69.2
	UniMiB	2	97.3	28.6	81.5	26.2	96.8	96.7	100.0	97.8	58.5	92.9	92.7	94.5
Gender	Motion sense	2	63.2	36.8	56.2	49.6	60.8	59.7	97.0	62.5	52.5	59.2	56.6	71.2
Position	mHealth	3	90.0	19.7	37.3	42.8	86.9	87.1	98.3	89.6	36.1	82.3	84.8	92.1
Detailed Act.	UniMiB	17	68.7	0.8	22.5	1.8	63.4	61.6	99.5	63.5	1.8	35.8	35.1	44.7

Using VGG10, identical to the anonymisation network, person-identification F1 scores drop below chance across all datasets, verifying effective identity removal; raw signals had yielded 50%–80% accuracy. Activity-recognition performance, however, reveals clear differences among methods. Traditional statistical transformations, RP [40] and NOS2R2 Gaussian noise [36], and the auto-encoder AAE [12] markedly degrade recognition accuracy. Each of these techniques induces a substantial distributional shift: RP rotates and compresses feature geometry, NOS2R2 injects broadband noise, and AAE generates a new latent space that assumes subsequent retraining on anonymised data. Models fitted to the original distribution therefore fail to extract useful patterns, producing large accuracy losses unless costly retraining is performed.

AAP, by contrast, adds minimal task-aware noise and preserves the structure on which existing classifiers rely. Across datasets, activity-recognition accuracy with AAP remains within 0.5–11% of the baseline while identity prediction stays at random level. On the UniMiB dataset, for example, detailed-activity classification falls to 1.8% with AAE but still reaches 63.4% under AAP. These results demonstrate that AAP achieves a favourable privacy–utility compromise: identity cues are suppressed, yet behaviourally relevant information is largely retained, and no model retraining is required.

Next, to evaluate robustness against changes in model architecture, the results obtained with ResNet10 are considered. Similar to VGG10, the person-identification accuracy consistently falls below chance level across all datasets. However, it was observed that the perturbation magnitude parameter α required for AAP tends to be higher compared than in the case with VGG10. In terms of activity recognition accuracy, trends similar to those observed with VGG10 were evident, with AAP consistently outperforming AAE. Nonetheless, due to the increased perturbation magnitude, performance with ResNet10 slightly decreased compared to that obtained with VGG10. This trend is similarly observed in other labels. Therefore, the proposed method demonstrates robustness against variations in model architecture, although increased perturbations slightly degrade information retention performance. The proposed perturbations are optimized on a per-dataset basis. Future work will explore domain-adversarial objectives and meta-learning schemes to improve cross-dataset transfer.

5.2.3 Evaluation of Waveform Preservation

In addition to evaluating classification performance, waveform preservation is assessed by comparing the MSE values (Table 3) and visualizing waveform changes (Fig. 5), both performed on time-domain data. The MSE values indicate that AAP produces considerably lower fluctuations than AAE across all datasets, which supports the improved classification performance observed earlier. Similarly, the visual results confirm that the variation from the original waveform is minimal. These results demonstrate that the proposed method effectively disrupts person recognition by classifiers while preserving the essential characteristics of the original signal.

Table 3: MSE between the original and anonymized signals for VGG10 model in time-domain data $(m/s^2)^2$

	AAE	AAP	F-AAP	MF-AAP
Motion sense	0.7587	0.0443	0.0486	0.0979
mHealth	0.7761	0.0206	0.0314	0.0511
UniMiB	0.9106	0.0194	0.0317	0.0544

In summary, the experimental results indicate that AAP not only effectively degrades person-identification accuracy but also preserves waveform characteristics, allowing models trained on raw data to be used directly without retraining. This leads to a more versatile anonymization approach compared to AAE.

5.3 RQ2: Effectiveness in Frequency Domain

Next, anonymization performance in the frequency domain is discussed. To the best of current knowledge, existing anonymization studies for activity recognition have not considered scenarios in which an attacker converts sensor waveforms into the frequency domain to conduct person-identification attacks. Thus, the present study investigates person-identification accuracy and activity-recognition accuracy when time-series sensor waveforms are transformed into the frequency domain using STFT.

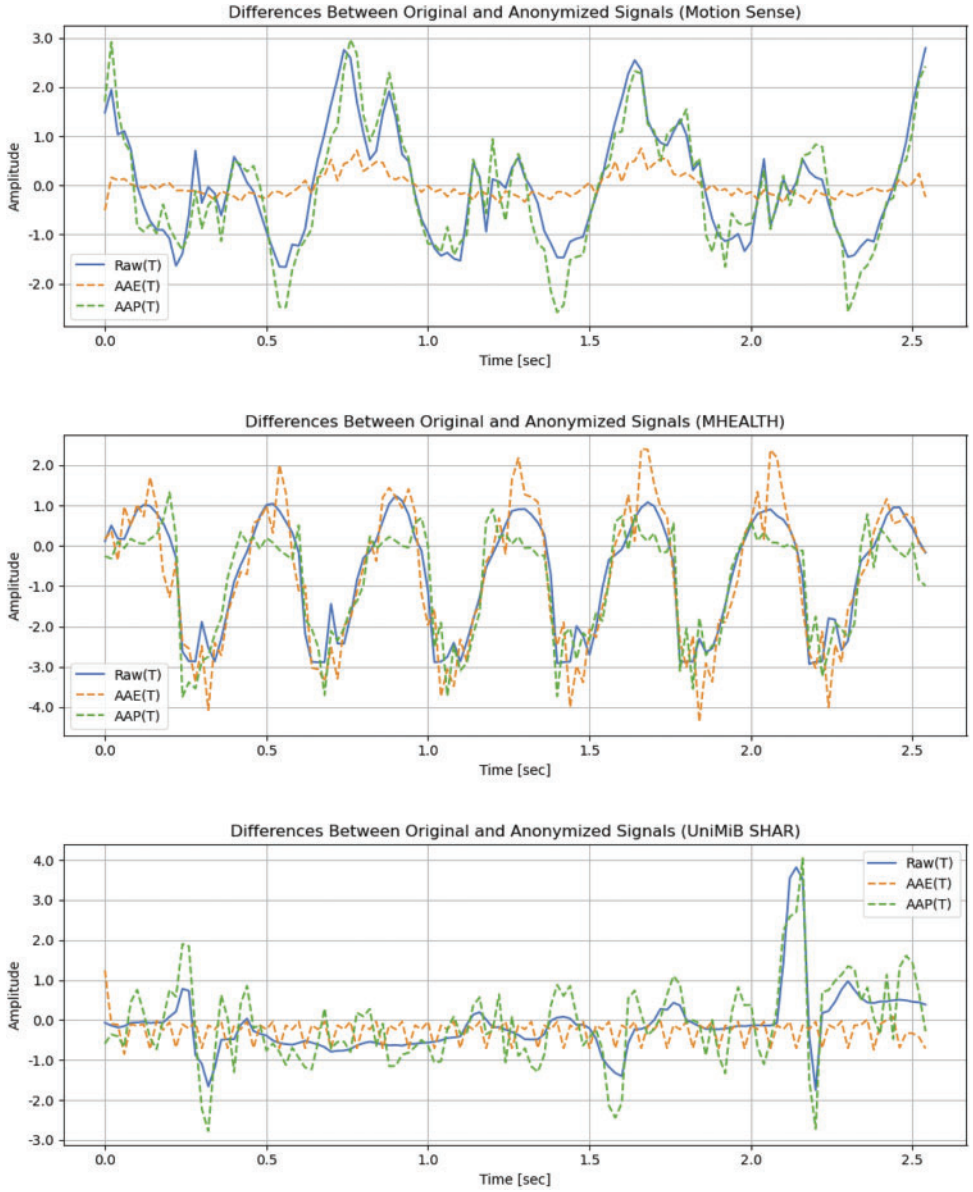

Figure 5: Waveform changes before and after anonymization

The experimental results are presented in Table 4. Experimental conditions are identical to those described in the previous section, except that the attacker's model targets frequency-domain data, requiring preprocessing to convert time-series sensor waveforms into the frequency domain before inputting them into the model. Examining the person-identification accuracy, it is observed that performance for all methods remains around the chance level. However, because AAP considers only the time domain, achieving anonymization below the chance level requires significantly increasing the perturbation magnitude parameter α. Consequently, although the performance of AAP remains superior to AAE, accuracy in activity recognition and other estimation tasks decreases compared to the results shown in Table 2. In contrast, F-AAP, which explicitly accounts for the frequency domain, maintains strong anonymization performance even with smaller α values, thus effectively preserving accuracy in activity recognition and other tasks.

Nevertheless, it was also found that the effectiveness of F-AAP diminishes when the model architecture changes, such as in the case of ResNet10.

Table 4: Comparison of anonymization performance on frequency-domain data [%]. In this scenario, the attacker has each model trained by frequency-domain sensor data

Model architecture			VGG 10					ResNet 10				
Test data			Raw	AAE	AAP	F-AAP	MF-AAP	Raw	AAE	AAP	F-AAP	MF-AAP
Target	Dataset	Class										
Person	Motion sense	24	54.9	0.3	4.1	3.4	3.9	55.8	0.7	3.9	3.9	3.9
	mHealth	10	83.6	5.4	9.2	4.8	6.2	79.6	4.5	9.7	9.7	0.0
	UniMiB	30	70.8	0.2	3.2	1.6	2.9	67.9	0.2	3.2	3.1	3.3
Activity	Motion sense	6	85.8	4.7	23.2	83.9	93.2	84.3	7.4	31.7	41.8	48.7
	mHealth	12	74.9	6.6	38.4	67.9	96.4	79.3	5.5	47.8	49.6	56.7
	UniMiB	2	99.2	37.1	91.9	98.8	99.9	99.3	28.6	91.2	93.4	95.6
Gender	Motion sense	2	63.2	41.8	54.3	58.4	96.5	60.4	35.0	55.8	55.2	74.5
Position	mHealth	3	89.0	22.0	77.0	88.6	99.1	89.9	21.4	80.0	80.4	85.2
Detailed Act.	UniMiB	17	63.5	1.2	14.3	51.0	97.6	62.4	0.7	12.3	26.3	35.5

These observations indicate that F-AAP effectively maintains anonymization performance against person-identification attacks in the frequency domain while mitigating adverse impacts on the accuracy of activity recognition and other tasks. However, robustness to differences in model architecture remains a challenge that should be addressed in future work.

5.4 RQ3: Selectively Suppress Person Identification Accuracy

The previously evaluated methods, AAP and F-AAP, require only the person labels during anonymization. Although these methods are easy to implement, their capability for information retention is limited. In contrast, the proposed method, MF-AAP, requires additional labels (e.g., activity, gender, etc.) in anonymization. However, this requirement enables selective control over information retention or suppression. Thus, this experiment investigates the effects of introducing perturbations designed to suppress person-identification performance while simultaneously enhancing the recognition performance of other attributes.

The results of this investigation are presented under the MF-AAP columns in Tables 2 and 4. Regardless of the model architecture or the target domain, the performance of person identification could be consistently suppressed below the chance level through adjustments of the parameter α. Furthermore, examination of activity recognition and other estimation accuracies shows performance improvements across all conditions. Notably, this improvement is particularly prominent for VGG10, which is the anonymization model itself, where most attributes achieved recognition accuracies exceeding 90%. Therefore, the results clearly demonstrate that MF-AAP can intentionally control both the enhancement and suppression of classification performance.

This phenomenon can be interpreted as embedding label information into the original data. AAP and F-AAP achieve anonymization by estimating person labels from input data and adding perturbations in the opposite direction of the gradient calculated with respect to those inputs. Although activity recognition and other labels are not explicitly utilized in the anonymization process, information retention is pursued by minimizing the perturbation magnitude. Conversely, MF-AAP employs multiple labels during anonymization and introduces perturbations using both forward and inverse gradients. Adding forward gradients to

the input introduces slight perturbations that improve classification performance, effectively embedding label-specific features into the original data.

6 Discussion

6.1 Effects of α

As a key feature of AAP, the intensity of the perturbation (α) can be adjusted to control the degree of anonymization (Int. adj.). In the experiments thus far, α is tuned so that person identification performance is forced to fall below the chance level. This section investigates how increasing or decreasing α affects overall performance. The following score is introduced as a new metric:

$$S = \frac{s_{\text{aap}} - c}{s_{\text{raw}} - c} \tag{4}$$

Here, S is calculated for each estimation target (Person, Activity, and others) based on its chance rate c, representing the relative ratio to the score s_{raw} achieved using raw data as the upper bound. Because a higher value is more desirable, $S'_{\text{person}} = 1.0 - S_{\text{person}}$ is defined only for the person-identification task. Finally, each score is clamped between 0.0 and 1.0.

Fig. 6 presents how the scores of each method change with different α values. For comparison, AAE is also plotted; however, because AAE does not allow adjusting the anonymization level after training, its value remains constant. Since the model architecture used here is ResNet10, these results illustrate the case where the evaluation model (ResNet10) differs from the model architecture employed in the anonymization (VGG10). From the figure, it is evident that, while AAE serves as a baseline with strong anonymization performance, it substantially degrades the estimation accuracy for Activity and Other, indicating significant information loss. In contrast, the proposed methods show that by setting α in the range of approximately 0.2–0.7, one can achieve comparable anonymization performance to AAE while still preserving performance on the other tasks.

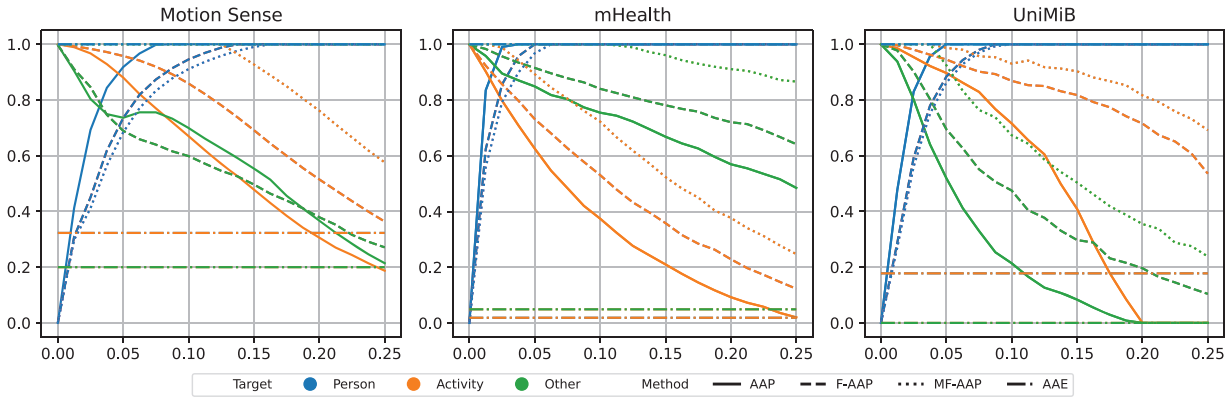

Figure 6: Comparison of target estimation score using ResNet10 in time-domain data across the mHealth, Motion Sense, and UniMiB datasets. Colors denote the different targets, and line styles indicate the applied methods (AAE, AAP, F-AAP, and MF-AAP)

When comparing the characteristics among the methods, AAP, F-AAP, and MF-AAP respond to changes in α in descending order of sensitivity. For instance, in the Motion Sense dataset, person identification accuracy saturates around $\alpha = 0.07$ for AAP, $\alpha = 0.13$ for F-AAP, and $\alpha = 0.16$ for MF-AAP. In each

case, the anonymization performance is comparable to that of AAE, and in many instances, even when α is further increased, these methods still retain more information than AAE.

6.2 Task-Interdependencies Analysis

Based on the preceding results, interdependencies among the subtasks are examined. First, the F1 scores obtained with Raw and AAP inputs in Table 4 are compared. Relevant rows are extracted and reorganized in Table 5, which lists the absolute drop ΔF_1 and the relative reduction RR = $\Delta F_1 / F_{1,\mathrm{Raw}}$. In the Motion Sense dataset, Activity suffers a substantial decrease of 62.6%, whereas Gender declines by only 8.9%. A similar pattern appears for Activity versus Position in mHealth. Conversely, on UniMiB the degradation in coarse activity recognition is minor (−7.3%), while fine–grained actions (Detailed Act.) deteriorate by 49.2%.

Table 5: Comparison of performance degradation on frequency-domain data (values in %). Based on Table 4. ΔF_1 is the F1 score drop from Raw to AAP; RR is the relative reduction, $\Delta F_1 / F_{1,Raw}$ (%). MI denotes the mutual information (bit) between the predicted class labels obtained on the Raw data

		Activity				Another				MI
		Raw	**AAP**	**ΔF_1**	**RR**	**Raw**	**AAP**	**ΔF_1**	**RR**	**[bit]**
Motion sense	Activity vs. Gender	85.8	23.2	−62.6	−73.0	63.2	54.3	−8.9	−14.1	0.010
mHealth	Activity vs. Position	74.9	38.4	−36.5	−48.7	89.0	77.0	−12.0	−13.5	0.044
UniMiB	Activity vs. Detailed Act.	99.2	91.9	−7.3	−7.4	63.5	14.3	−49.2	−77.5	0.543

Next, the findings are interpreted through mutual information (MI). The MI column in Table 5 reports the bit-wise MI between the predicted labels of each task on the unperturbed (Raw) data. The MI for the Activity–Gender and Activity–Position pairs is very low (0.01–0.044 bit), indicating near-independence. In contrast, the Activity–Detailed Act. pair exhibits a high MI of 0.543 bit, revealing strong information overlap. These quantitative results align with intuition: coarse activity labels share little information with gender or sensor placement, whereas fine-grained action classes inherently overlap with the broader activity categories.

6.3 Computational Efficiency

First, we examine the architectural differences among the anonymization methods.

- **AAE** consists of an encoder-decoder pair, two identity-recognition subnetworks (one attached to the encoder, one to the decoder), and one activity recognition subnetwork.
- The **plain AAP** variant contains only one identity-recognition network.
- **F-AAP** employs two identity-recognition networks, one for the time-domain input and one for the frequency-domain input.
- **MF-AAP** holds twice as many subnetworks as the number of downstream tasks.

Next, we consider the computational steps required at inference time (i.e., during anonymization).

- For **AAE**, a single forward pass through the encoder and then the decoder is sufficient.
- **AAP** must execute a forward pass and a backward pass through its identity model for every iteration of the perturbation update.
- **F-AAP** adds forward-backward passes in both the time and frequency domains plus the cost of STFT and inverse STFT.
- In **MF-AAP**, the number of model inferences scales with the number of tasks.

These observations are summarized in Table 6. Assuming that a backward pass costs roughly twice as many FLOPs as a forward pass, the theoretical requirements become AAE: 81.5 MFLOPs, AAP: 99.5 t MFLOPs, where t denotes the number of iterations. The Latency [s] row in Table 6 reports the actual anonymization time measured on the mHealth test set. All experiments were run on a machine equipped with an Intel Core i9–13900KF, 32 GB RAM, and an NVIDIA RTX 4090 GPU.

Table 6: Comparison of computational costs for each anonymization method

Metric	AAE	AAP	F-AAP	MF-AAP
# iterations	1	$t = 15$	$t = 15$	$t = 15$
MFLOPs	81.5	99.5 t	–	–
Latency [s]	0.09	0.38	15.97	16.78

From the table we observe that, although AAP takes about four times longer than AAE, the gap is far smaller than the FLOP counts alone would suggest. Conversely, the running time of F-AAP and MF-AAP is dominated by the additional STFT/ISTFT transforms, leading to a substantial increase in total computation time.

6.4 Limitations

6.4.1 Scenario in Which the Anonymization Method and Raw Data Are Leaked

In the previous section, the privacy risk under scenario (b), in which a personal identification model and scenario (a) application-server login credentials were compromised, was evaluated, and it was demonstrated that the proposed anonymization approach could adequately preserve useful information while protecting privacy (Section 3.2). This section further examines scenario (c), in which raw sensor data and personal IDs are also leaked. When only scenario (c) is compromised, the attacker can implement a personal identification model trained on raw data, resulting in a risk level similar to scenario (b). However, if the anonymization method is independently leaked from another source, the attacker may be able to reconstruct the anonymized sensor data.

To quantify this risk, performance in a setting where the attacker has access to a portion of labeled anonymized sensor data is compared. Table 7 presents the results, indicating that if a model trained on anonymized data is also exposed, none of the proposed methods can effectively maintain anonymization. Moreover, person identification becomes even easier than with the raw data alone, primarily because the proposed anonymization leverages adversarial training [52]. While AAE also allows for some degree of person identification under these circumstances, it does so to a lesser extent than the methods. Therefore, these findings suggest that developing anonymization techniques robust against leaks of anonymized waveforms remains an important open challenge.

Table 7: Comparison of anonymization performance on time-domain sensor data when the anonymized dataset were leaked

Model architecture			VGG10			ResNet10		
Train & test data			AAE	AAP	MF-AAP	AAE	AAP	MF-AAP
Target	Dataset	Class						
Person	Motion sense	24	48.6	81.7	84.2	47.6	81.1	82.8
	mHealth	10	57.8	93.6	93.9	60.7	88.0	89.2
	UniMiB	30	50.6	88.5	74.7	47.1	83.2	69.6
Activity	Motion sense	6	96.8	81.3	95.1	96.7	81.7	94.2
	mHealth	12	81.2	70.8	100.0	80.8	80.2	100.0
	UniMiB	2	96.3	96.6	98.9	96.5	96.1	97.8
Gender	Motion sense	2	66.0	65.7	82.3	67.3	64.2	82.9
Position	mHealth	3	87.3	89.0	100.0	86.6	88.4	100.0
Detailed Act.	UniMiB	17	52.0	57.4	79.7	51.1	50.4	63.4

6.4.2 Other Attacks

Since the proposed method conceals information by adding the inverse gradient of the person-identification loss to sensor data, it is primarily tailored to obscuring features associated with a known label (e.g., person identity). Consequently, it may be vulnerable to side-channel attacks that can infer sensitive information without explicitly identifying the individual. For example, in keystroke inference [56], an attacker might extract touchscreen inputs from motion-sensor data. The proposed approach would require explicit label information for these keystrokes (i.e., keys pressed) to compute the inverse gradient for anonymization, which becomes prohibitively expensive to implement in practice. Moreover, if a method infers user location from motion-sensor data without using GPS information [57], effective label assignment could be infeasible. In such cases, the proposed approach cannot be applied, underscoring a broader limitation when label annotation is either incomplete or difficult to obtain.

7 Conclusion

This study presented a novel sensor-data anonymization framework leveraging AP for privacy protection in wearable sensing applications. Unlike autoencoder-based methods such as the AAE, the proposed approach adds minimal, targeted noise to raw waveforms, thereby suppressing person-identification accuracy while retaining greater task-relevant information. Frequency-informed (F-AAP) and multi-task (MF-AAP) extensions are further introduced to address threats from frequency-domain analysis and to selectively preserve specific classification tasks (e.g., activity recognition and gender estimation). Extensive evaluations on three public datasets demonstrated that the proposed methods can degrade identification performance to near-chance levels even against unseen model architectures, while substantially preserving or enhancing performance on other tasks. Moreover, experimental results indicate that critical waveform characteristics remain relatively intact, facilitating the reuse of established downstream models trained on raw data. These findings suggest that AP-based anonymization offers a compelling and flexible alternative to conventional approaches, meeting the increasing need for effective privacy protection without sacrificing diverse analytical utility.

Acknowledgement: The authors are grateful to all the editors and anonymous reviewers for their comments and suggestions. This manuscript's English translation and proofreading were assisted by ChatGPT, a large language model developed by OpenAI. All responsibility for the content of this paper rests solely with the authors.

Funding Statement: This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant-in-Aid for Scientific Research (C) under Grants 23K11164.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Tatsuhito Hasegawa; methodology, Kyosuke Fujino; software, Kyosuke Fujino; validation, Tatsuhito Hasegawa and Kyosuke Fujino; formal analysis, Tatsuhito Hasegawa and Kyosuke Fujino; investigation, Kyosuke Fujino; resources, Tatsuhito Hasegawa; data curation, Kyosuke Fujino; writing—original draft preparation, Kyosuke Fujino; writing—review and editing, Tatsuhito Hasegawa; visualization, Tatsuhito Hasegawa and Kyosuke Fujino; supervision, Tatsuhito Hasegawa; project administration, Tatsuhito Hasegawa; funding acquisition, Tatsuhito Hasegawa. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are openly available in following repositories: Motion Sense Dataset [15]: https://github.com/mmalekzadeh/motion-sense (accessed on 28 May 2025). mHealth [16]: https://archive.ics.uci.edu/dataset/319/mhealth+dataset (accessed on 28 May 2025). UniMiB SHAR Dataset [17]: http://www.sal.disco.unimib.it/technologies/unimib-shar/ (accessed on 28 May 2025).

Ethics Approval: This study did not require ethics approval as it only used publicly available anonymized datasets, and no new data were collected from human subjects.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

Abbreviations

AAE	Anonymizing autoencoder
AP	Adversarial perturbations
AAP	Anonimizing adversarial perturbations
F-AAP	Frequency-informed AAP
MF-AAP	Multi-task frequency-informed AAP
HAR	Human activity recognition
RP	Random projection
GAN	Generative adversarial networks
PGAN	Private GAN
MSE	Mean squared error
RQ	Research question
STFT	Short-time Fourier transform

References

1. Mamdiwar SD, R. A, Shakruwala Z, Chadha U, Srinivasan K, Chang CY. Recent advances on IoT-assisted wearable sensor systems for healthcare monitoring. Biosensors. 2021;11(10):372. doi:10.3390/bios11100372.
2. Li C, Bian Y, Zhao Z, Liu Y, Guo Y. Advances in biointegrated wearable and implantable optoelectronic devices for cardiac healthcare. Cyb Bionic Syst. 2024;5(33):0172. doi:10.34133/cbsystems.0172.
3. Xing Y, Yang K, Lu A, Mackie K, Guo F. Sensors and devices guided by artificial intelligence for personalized pain medicine. Cyb Bionic Syst. 2024;5:0160. doi:10.34133/cbsystems.0160.
4. Lam Po Tang S. 8—Wearable sensors for sports performance. In: Shishoo R, editor. Textiles for sportswear. Woodhead publishing series in textiles. Sawston, UK: Woodhead Publishing; 2015. p. 169–96. doi:10.1016/B978-1-78242-229-7.00008-4.

5. Mencarini E, Rapp A, Tirabeni L, Zancanaro M. Designing wearable systems for sports: a review of trends and opportunities in human-computer interaction. IEEE Trans Hum Mach Syst. 2019;49(4):314–25. doi:10.1109/thms.2019.2919702.
6. Motti VG, Caine K. Users' privacy concerns about wearables: impact of form factor, sensors and type of data collected. In: Financial Cryptography and Data Security: FC 2015. 1st ed. Berlin/Heidelberg, Germany: Springer; 2015. p. 231–44 doi:10.1007/978-3-662-48051-9_17.
7. Xu Z, He D, Vijayakumar P, Gupta BB, Shen J. Certificateless public auditing scheme with data privacy and dynamics in group user model of cloud-assisted medical WSNs. IEEE J Biomed Health Inform. 2023;27(5):2334–44. doi:10.1109/jbhi.2021.3128775.
8. Huhn S, Axt M, Gunga HC, Maggioni MA, Munga S, Obor D, et al. The impact of wearable technologies in health research: scoping review. JMIR Mhealth Uhealth. 2022;10(1):e34384.
9. Sarker H, Sharmin M, Ali AA, Rahman MM, Bari R, Hossain SM, et al. Assessing the availability of users to engage in just-in-time intervention in the natural environment. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2014 Sep 13–17; Seattle, DC, USA. p. 909–20.
10. Fujimoto R, Nakamura Y, Arakawa Y. Differential privacy with weighted for privacy-preservation in human activity recognition. In: 2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, PerCom Workshops 2023; Atlanta, GA, USA. 2023. p. 634–9.
11. Menasria S, Lu M, Dahou A. PGAN framework for synthesizing sensor data privately. J Inf Secur Appl. 2022;67(2):103204. doi:10.1016/j.jisa.2022.103204.
12. Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H. Mobile sensor data anonymization. In: Proceedings of the International Conference on Internet of Things Design and Implementation. IoTDI '19. ACM; Montreal, QC, Canada; 2019 Apr 15–18; p. 49– 58. doi:10.1145/3302505.3310068.
13. Bigelli L, Contoli C, Freschi V, Lattanzi E. Privacy preservation in sensor-based human activity recognition through autoencoders for low-power IoT devices. Internet of Things. 2024;26(6):101189. doi:10.1016/j.iot.2024.101189.
14. Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, et al. Intriguing properties of neural networks. arXiv:1312.6199. 2013.
15. Malekzadeh M, Clegg RG, Cavallaro A, Haddadi H. Protecting sensory data against sensitive inferences. In: Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems; 2018 Apr 23–26; Porto, Portugal. p. 1–6. doi:10.1145/3195258.3195260.
16. Banos O, Garcia R, Holgado-Terriza JA, Damas M, Pomares H, Rojas I, et al. mHealthDroid: a novel framework for agile development of mobile health applications. In: Pecchia L, Chen LL, Nugent C, Bravo J, editors. Ambient assisted living and daily activities. Cham, Switzerland: Springer International Publishing; 2014. p. 91–8. doi:10.1007/978-3-319-13105-4_14.
17. Micucci D, Mobilio M, Napoletano P. UniMiB SHAR: a dataset for human activity recognition using acceleration data from smartphones. Appl Sci. 2017;7(10):1101. doi:10.3390/app7101101.
18. Xu X, Han M, Nagarajan SM, Anandhan P. Industrial Internet of Things for smart manufacturing applications using hierarchical trustful resource assignment. Comput Commun. 2020;160(2):423–30. doi:10.1016/j.comcom.2020.06.004.
19. Talebkhah M, Sali A, Marjani M, Gordan M, Hashim SJ, Rokhani FZ. IoT and big data applications in smart cities: recent advances, challenges, and critical issues. IEEE Access. 2021;9:55465–84. doi:10.1109/access.2021.3070905.
20. Lee MW, Khan AM, Kim TS. A single tri-axial accelerometer-based real-time personal life log system capable of human activity recognition and exercise information generation. Personal Ubiquitous Comput. 2011;15(8):887–98. doi:10.1007/s00779-011-0403-3.
21. Xu H, Pan Y, Li J, Nie L, Xu X. Activity recognition method for home-based elderly care service based on random forest and activity similarity. IEEE Access. 2019;7:16217–25. doi:10.1109/access.2019.2894184.
22. Inoue S, Ueda N, Nohara Y, Nakashima N. Recognizing and understanding nursing activities for a whole day with a big dataset. J Inf Process. 2016;24(6):853–66. doi:10.2197/ipsjjip.24.853.
23. Kwapisz JR, Weiss GM, Moore S. Activity recognition using cell phone accelerometers. SIGKDD Explor. 2011;12(2):74–82. doi:10.1145/1964897.1964918.

24. Shoaib M, Bosch S, Incel ÖD, Scholten H, Havinga PJM. Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors. 2016;16(4):426. doi:10.3390/s16040426.
25. Voicu RA, Dobre C, Bajenaru L, Ciobanu RI. Human physical activity recognition using smartphone sensors. Sensors. 2019;19(3):458. doi:10.3390/s19030458.
26. Han F, Yang P, Du H, Li XY. *Accuth*$^{+}$+: accelerometer-based anti-spoofing voice authentication on wrist-worn wearables. IEEE Trans Mob Comput. 2024;23(5):5571–88. doi:10.1109/tmc.2023.3314837.
27. Li F, Shirahama K, Nisar MA, Köping L, Grzegorzek M. Comparison of feature learning methods for human activity recognition using wearable sensors. Sensors. 2018;18(2):679. doi:10.3390/s18020679.
28. Mehmood K, Imran HA, Latif U. HARDenseNet: A 1D densenet inspired convolutional neural network for human activity recognition with inertial sensors. In: 2020 IEEE 23rd International Multitopic Conference (INMIC); 2020 Nov 5–7; Bahawalpur, Pakistan. p. 1–6.
29. Ronald M, Poulose A, Han DS. iSPLInception: an inception-resnet deep learning architecture for human activity recognition. IEEE Access. 2021;9:68985–9001. doi:10.1109/access.2021.3078184.
30. Al-qaness MAA, Dahou A, Abd Elaziz M, Helmi AM. Human activity recognition and fall detection using convolutional neural network and transformer-based architecture. Biomed Signal Process Control. 2024;95(3):106412. doi:10.1016/j.bspc.2024.106412.
31. Hussain A, Khan SU, Khan N, Bhatt MW, Farouk A, Bhola J, et al. A hybrid transformer framework for efficient activity recognition using consumer electronics. IEEE Trans Consum Electron. 2024;70(4):6800–7. doi:10.1109/tce.2024.3373824.
32. Pareek G, Nigam S, Singh R. Modeling transformer architecture with attention layer for human activity recognition. Neural Comput Appl. 2024;36(10):5515–28. doi:10.1007/s00521-023-09362-7.
33. Neves F, Souza R, Sousa J, Bonfim M, Garcia V. Data privacy in the Internet of Things based on anonymization: a review. J Comput Secur. 2023;31(3):261–91. doi:10.3233/JCS-210089.
34. Yang Y, Hu P, Shen J, Cheng H, An Z, Liu X. Privacy-preserving human activity sensing: a survey. High-Confid Comput. 2024;4(1):100204. doi:10.1016/j.hcc.2024.100204.
35. Debs N, Jourdan T, Moukadem A, Boutet A, Frindel C. Motion sensor data anonymization by time-frequency filtering. In: 2020 28th European Signal Processing Conference (EUSIPCO); 2021 Jan 18–21; Amsterdam, Netherlands. p. 1707–11.
36. Rahman M, Paul MK, Sattar AHMS. Efficient perturbation techniques for preserving privacy of multivariate sensitive data. Array. 2023;20(4):100324. doi:10.1016/j.array.2023.100324.
37. Shou L, Shang X, Chen K, Chen G, Zhang C. Supporting pattern-preserving anonymization for time-series data. IEEE Trans Knowl Data Eng. 2013;25(4):877–92. doi:10.1109/tkde.2011.249.
38. Samarati P, Sweeney L. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. In: Proceedings of the 1998 IEEE Symposium on Research in Security and Privacy (S&P). 1998 May 3–6; Oakland, CA, USA.
39. Liu F, Li T. A clustering K-anonymity privacy-preserving method for wearable IoT devices. Secur Commun Netw. 2018;2018(1):4945152.
40. Liu K, Kargupta H, Ryan J. Random projection-based multiplicative data perturbation for privacy preserving distributed data mining. IEEE Trans Knowl Data Eng. 2006;18(1):92–106. doi:10.1109/tkde.2006.14.
41. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S et al. Generative adversarial nets. In: Advances in neural information processing systems. Red Hook, NY, USA: Curran Associates, Inc.; 2014. Vol. 27. doi:10.1145/3422622.
42. Hallyburton T, Nair NR, Moya Rueda F, Grzeszick R, Fink GA. Anonymisation for time-series human activity data. In: Pattern recognition. Cham, Switzerland: Springer Nature; 2025. p. 17–32. doi:10.1007/978-3-031-78354-8_2.
43. Aleroud A, Shariah M, Malkawi R, Khamaiseh SY, Al-Alaj A. A privacy-enhanced human activity recognition using GAN & entropy ranking of microaggregated data. Cluster Comput. 2024;27(2):2117–32.

44. Iwasawa Y, Nakayama K, Yairi I, Matsuo Y. Privacy issues regarding the application of DNNs to activity-recognition using wearables and its countermeasures by use of adversarial training. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17; 2017 Aug 19–25; Melbourne, VIC, Australia. p. 1930–6. doi:10.24963/ijcai.2017/268.
45. Boutet A, Frindel C, Gambs S, Jourdan T, Ngueveu RC. DySan: dynamically sanitizing motion sensor data against sensitive inferences through adversarial networks. In: Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security. ASIA CCS '21. 2021 Jun 7–11; Hong Kong, China. p. 672–86. doi:10.1145/3433210.3453095.
46. Wang G, Guo H, Wang Y, Chen B, Zhou C, Yan Q. Protecting activity sensing data privacy using hierarchical information dissociation. In: 2024 IEEE Conference on Communications and Network Security (CNS); 2024 Sep 30–Oct 3; Taipei, Taiwan. p. 1–9.
47. Ahmad S, Morerio P, Del Bue A. Person re-identification without identification via event anonymization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2023 Oct 1–6; Paris, France. p. 11132–41.
48. Zhang D, Yao L, Chen K, Yang Z, Gao X, Liu Y. Preventing sensitive information leakage from mobile sensor signals via integrative transformation. IEEE Trans Mob Comput. 2022;21(12):4517–28. doi:10.1109/tmc.2021.3078086.
49. Kurakin A, Goodfellow IJ, Bengio S. Adversarial machine learning at scale. In: International Conference on Learning Representations; 2017 Apr 24–26; Toulon, France.
50. Xie C, Zhang Z, Zhou Y, Bai S, Wang J, Ren Z, et al. Improving transferability of adversarial examples with input diversity. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 2725–34.
51. Dong Y, Pang T, Su H, Zhu J. Evading defenses to transferable adversarial examples by translation-invariant attacks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 4307–16.
52. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. arXiv:1412.6572. 2015.
53. Chung MK. Gaussian kernel smoothing. arXiv:2007.09539. 2021.
54. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015); 2015 May 7–9; San Diego, CA, USA. p. 1–14.
55. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016; Las Vegas, NV, USA. p. 770–8.
56. Zheng S, Shi P, Xu H, Zhang C. Launching the new profile on facebook: understanding the triggers and outcomes of users' privacy concerns. In: Trust and trustworthy computing; Berlin/Heidelberg, Germany: Springer; 2012. p. 325–39.
57. Narain S, Vo-Huu TD, Block K, Noubir G. Inferring user routes and locations using zero-permission mobile sensors. In: 2016 IEEE Symposium on Security and Privacy (SP); 2016 May 22–26; San Jose, CA, USA. p. 397–413.