

Privacy and Utility Preserving Sensor-Data Transformations¹

Mohammad Malekzadeh*, Richard G. Clegg*, Andrea Cavallaro*, Hamed Haddadi**
 {m.malekzadeh, r.clegg, a.cavallaro}@qmul.ac.uk, h.haddadi@imperial.ac.uk
 *Queen Mary University of London, **Imperial College London

Abstract

Sensitive inferences and user re-identification are major threats to privacy when raw sensor data from wearable or portable devices are shared with cloud-assisted applications. To mitigate these threats, we propose mechanisms to transform sensor data before sharing them with applications running on users' devices. These transformations aim at eliminating patterns that can be used for user re-identification or for inferring potentially sensitive activities, while introducing a minor utility loss for the target application (or task). We show that, on gesture and activity recognition tasks, we can prevent inference of potentially sensitive activities while keeping the reduction in recognition accuracy of non-sensitive activities to less than 5 percentage points. We also show that we can reduce the accuracy of user re-identification and of the potential inference of gender to the level of a random guess, while keeping the accuracy of activity recognition comparable to that obtained on the original data.

1. Introduction

Sensors such as accelerometer, gyroscope, and magnetometer, embedded in personal smart devices generate data that can be used to monitor users' activities, interactions, and mood [1, 2, 3]. Applications (apps) installed on smart devices can get access to raw sensor data to make *required* (*i.e.* desired) inferences for tasks such as gesture or activity recognition. However, sensor data can also facilitate some potentially *sensitive* (*i.e.* undesired) inferences that a user might wish to keep private, such as discovering smoking habits [4] or revealing personal attributes such as age and gender [5]. Some patterns in raw sensor data may also enable user re-identification [6].

Information privacy can be defined as “*the right to select what personal information about me is known to what people*” [7]. To preserve privacy, we need mechanisms to control the type and amount of information that providers of cloud-assisted apps can discover from sensor data. The main objective is to move

¹Accepted to appear in Pervasive and Mobile computing (PMC) Journal, Elsevier.

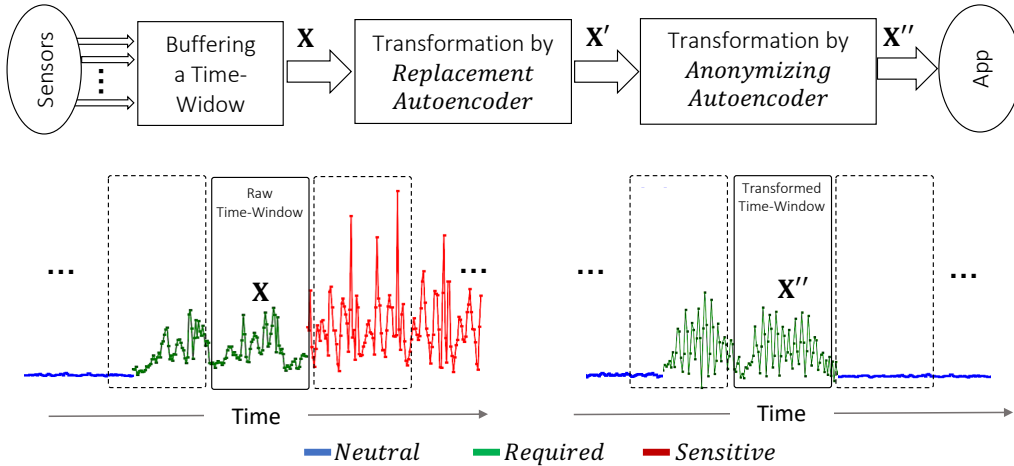


Figure 1: (Top) the data flows in the compound framework. At the test-time, first RAE automatically replaces *sensitive* time-windows with non-sensitive *neutral* data, while *required* time-windows are passed with minimal distortions. Then, AEE transforms data to reduce the chance of *user re-identification*. (Bottom) visual illustration of our transformation mechanism. Depicted signals show accelerometer data transformation for standing, walking, and jogging activities respectively as *neutral*, *required*, and *sensitive* inferences (from experiment of Section 4.3).

from the current binary setting of granting or not sensor permission to an app, toward a model that allows users to grant each app permission over a controlled range of inferences according to the target task. The challenging task is to design a mechanism with an acceptable trade-off between the protection of *sensitive* information and the maintenance of the *required* information for an inference [8]. To this end, we use *neutral* inferences that are irrelevant to the target task and not critical to the user’s privacy.

As specific example of categorization of *sensitive*, *required*, and *neutral* information, let us consider a smartwatch step-counter app: *required* information is essential for the app’s utility, such as walking or stair stepping; *sensitive* information is about activities a user wishes to keep private, such as smoking or typing on a keyboard, or information such as gender; whereas *neutral* information leads to inferences that are neither *required* nor *sensitive*, such as when the user sits or stands. Note that two types of information (*i.e.* *required* and *sensitive*, or *neutral* and *sensitive*) are sometimes entangled in the data of the same temporal window of sensor measurements. While locally differentially private mechanisms [9] provide plausible deniability guarantees when estimating, for example, the mean or frequency of a common variable among users [10], with multi-dimensional data released sequentially a more practical privacy model is *inferential privacy* [11, 12, 13] that measures the difference between an adversary’s belief about *sensitive* inferences before and after observing the released data.

We assume the app provider is honest in stating its *required* inferences that need to be made on the data, but it is also curious about making other unstated inferences that may reveal sensitive information and thus violate privacy. We

define *utility* as the accuracy in making the *required* inferences on the released data, and *privacy loss* as the accuracy in making *sensitive* inferences. We define the app provider as adversary and quantify the *privacy loss* as the improvement in the adversary’s posterior belief on making a *sensitive* inference by observing the data. Our proposed mechanisms aim to minimize the *privacy loss* while maintaining the *utility* of the raw data.

In this paper, we present mechanisms for transforming time-windows of sensor data to preserve privacy and utility during information disclosure [14, 15, 16] to a *honest-but-curious* app running on users’ devices. Specifically, we introduce a Replacement AutoEncoder (RAE) to protect *sensitive* inferences and an Anonymizing AutoEncoder (AAE) to prevent *user re-identification*, as well as a compound architecture by cascading the RAE and AAE (see Figure 1). The RAE and AAE can be deployed as interface into the devices’ operating system to enable users to choose whether to share their sensor data with an app directly or after transformations. To validate our mechanisms, in addition to using available datasets, we collected a dataset of activity recognition using smartphone sensors, which is made publicly available². Experiments on gesture and activity recognition show that the RAE substantially reduces the *privacy loss* for *sensitive* gestures or activities while limiting the reduction in the *utility* of the *required* and *neutral* gestures or activities to less than 5 percentage points. Furthermore, results on an activity recognition dataset of 24 users show a promising trade-off, with the utility maintained over 92% and a reduction of the *privacy loss* in user re-identification to less than 7% , from an initial 96% on the raw data. We also show that our mechanisms lead to models that can generalize across datasets and can be applied to new data of unseen users.

2. Related Work

Privacy-preserving mechanisms for time-series data can be implemented through *perturbations*, *synthesis*, *filtering*, or *transformations*.

Mechanisms using *perturbations* hide sensitive patterns by adding a crafted noise to each time-window of the time-series. The objective is to prevent perturbed data from including sufficient information to accurately reconstruct the original data [17]. Because an independent and identically distributed noise can be easily removed from correlated time-series [18], to reduce the risk of information leakage, the correlation between noise and original time-series should be indistinguishable [19]. For multi-dimensional sensor data, it is not easy to find a reliable model of correlation between the original data and an adequate noise. Hence, when general time-series perturbation approaches are extended to sensor data, effectively hiding sensitive patterns without excessively perturbing the non-sensitive ones is very challenging.

²Code and data are available at: <https://github.com/mmalekzadeh/motion-sense>

Table 1: Privacy-preserving mechanisms for sharing time series. Key - *Local*: applied on the user side (instead of being done globally by a data curator); *GesAct*: hides users’ sensitive gestures or activities; *Identity*: prevents user re-identification; *Sensors*: evaluated on sensor data; *Unseen*: can be used for data of users who did not contribute training data.

<i>Mechanism</i>	<i>Reference</i>	<i>Local</i>	<i>GesAct</i>	<i>Identity</i>	<i>Sensor</i>	<i>Unseen</i>
Perturbations	[17, 18, 19]		✓		✓	
Synthesis	[22, 24, 25]			✓		
Filtering	[26, 30, 33]	✓	✓			
	[27]		✓	✓	✓	
Transformations	[12, 13]	✓		✓		
	[11, 31]			✓	✓	
	[29]	✓		✓	✓	✓
Filtering & Transformations	Ours	✓	✓	✓	✓	✓

Data can also be *synthesized* to maintain some required statistics of the original data without information that can be used for re-identification. Adversarial learning enables one to approximate an underlying distribution to generate new data that are similar to the existing ones [20, 21]. To provide a privacy guarantee, generators can be trained under the constraint of differential privacy [22, 23] or with constraints on the type of information that should be unsynthesized in the data [24]. However, these mechanisms are used for offline dataset publishing by a data aggregator [25], not for online data transformation at the user side.

Filtering can be used to remove unwanted components only in temporal intervals that include sensitive information. MaskIt [26] releases location time-series when users are at a regular workplace and suppresses them when they are in a sensitive place, such as a hospital. A Markov chain built on a pre-defined set of conditions is employed for each user. A Dynamic Bayesian Network model can be used offline to replace sensitive time-windows that indicate users’ stress, while keeping non-sensitive time-windows corresponding to their walking periods [27].

Transformations can reduce the amount of sensitive information in the data by reconstruction [12] or by projecting each data sample into a lower dimensional latent representation [13, 28]. The information bottleneck in the hidden layers of neural networks helps to capture the main factors of variation in the data and to identify and obscure sensitive patterns in the latent representation [13], as well as during the reconstruction from the extracted low-dimensional representation [29, 30]. Global mechanisms involve a trusted data curator and, based on the information bottleneck principle, compress sensor data to reduce sensitive information that is irrelevant to the main task [11].

Table 1 compares methods related to our work. A privacy-preserving mechanism can be run globally or locally. Global mechanisms involve a trusted data curator that has access to the original data and offer a data transformation service to remove sensitive information before data publishing [11, 31, 32]. Local mech-

Table 2: Main notation used in this paper.

$\mathbf{x}_{sj} \in \mathbb{R}$	reading from sensor component s at sampling instant j ;
$\mathbf{X} \in \mathbb{R}^{M \times W}$	time-window of W samples from M sensors;
$\mathcal{X}^i, \mathcal{X}^o \subseteq \mathbb{R}^{M \times W}$	input and output datasets, respectively, for training the RAE;
$\mathbf{X}', \mathbf{X}'' \in \mathbb{R}^{M \times W}$	output of the RAE and the AAE, respectively;
$\mathbf{U} \in \{0, 1\}^N$	N -dim vector representing the identity of a user ($\sum_{i=1}^N u_i = 1$);
$\mathbf{Y} \in \{0, 1\}^B$	B -dim vector representing a gesture/activity ($\sum_{i=1}^B y_i = 1$);
$I(\cdot; \cdot)$	mutual information function;
$d(\cdot, \cdot)$	distance function between two time-series (<i>e.g.</i> Mean Squared Error).

anisms, instead, manipulate data at the user side, without relying on a trusted curator [26, 29, 30]. Our mechanisms run *locally* and can be used by users who did not contribute training data (*unseen* users).

3. Sensor-Data Transformation

We first introduce the Replacement AutoEncoder (RAE) that protects *sensitive* inferences, then we present the Anonymizing AutoEncoder (AAE) that prevents *user re-identification*. The notations we use in this paper are shown in Table 2.

3.1. Replacement AutoEncoder

Deep neural networks (DNNs) are powerful machine learning algorithms that progressively learn hierarchical and relevant representations of their training data. Earlier layers of a DNN can encode generic low-level data patterns and later layers can capture more specific high-level features. Autoencoders learn features from data through minimizing the differences (*e.g.* mean squared error or cross entropy) between the input and its reconstruction. The information bottleneck [34] in the hidden layers forces an autoencoder to put more attention on the descriptive data patterns in order to generalize the model.

Let a fixed-length time-window of sensor data, $\mathbf{X} = (\mathbf{x}_{sj}) \in \mathbb{R}^{M \times W}$, contain some specific patterns that are utilized to recognize the gesture or activity of the user at that specific time-point. For example, let us consider an smartwatch app which counts users’ daily steps. Users may want this app to only be able to infer activities that are *required* for step counting task, not other activities such as smoking or eating that may be considered *sensitive*. The main idea of RAE is to automatically recognize and replace each time-widows that reveals *sensitive* activities with a same dimension data that simulates a *neutral* activity, such as standing or sitting, which does not affect the step counter utility.

Let the training dataset include labeled sample time-windows, each belonging to one of the following categories: *required*, *sensitive*, or *neutral*. Let \mathcal{X}^i be

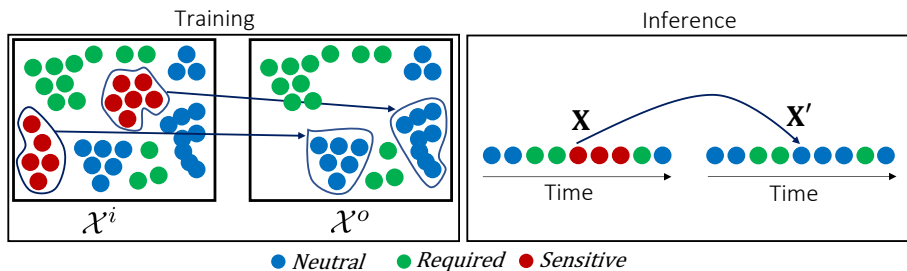


Figure 2: Circles represent time-windows in the input (\mathcal{X}^i) and output (\mathcal{X}^o) datasets for training the RAE. We first make a copy of the original input dataset and replace every *sensitive* time-window with a randomly chosen *neutral* one to prepare the output dataset for training the RAE. Then, the RAE is trained to transform each \mathbf{X}^i to the corresponding \mathbf{X}^o . At inference time, RAE can replace unseen *sensitive* time-windows with data that simulates *neutral* ones.

the *input* dataset and \mathcal{X}^o be the *output* dataset, with a one-to-one relationship between each $\mathbf{X}^i \in \mathcal{X}^i$ and an $\mathbf{X}^o \in \mathcal{X}^o$ explained in Figure 2. Basically, data samples of *sensitive* classes in \mathcal{X}^i are randomly replaced with data samples from one of the *neutral* classes to build \mathcal{X}^o . Therefore \mathcal{X}^o contains only samples from the *required* and *neutral* classes. The RAE is then trained to transform each \mathbf{X}^i to the corresponding \mathbf{X}^o , subject to a loss function, $L_R(\mathbf{X}^i, \mathbf{X}^o)$, which calculates the difference between the input of the RAE and its corresponding output.

Let a replacement be defined as *privacy-preserving* if its outcome removes, or practically bounds, the possibility of revealing *sensitive* inferences. If \mathbf{X}^r is a *privacy-preserving* replacement for *sensitive* data \mathbf{X} , the RAE aims to implement the following operation:

$$\mathbf{X}' = RAE(\mathbf{X}) = \begin{cases} \mathbf{X}^r & \text{if } \mathbf{X} \text{ reveals a sensitive inference,} \\ \mathbf{X} & \text{otherwise,} \end{cases} \quad (1)$$

where the *privacy loss* of the replaced data, \mathbf{X}' , is equivalent to the amount of sensitive information it reveals. If $R(\mathbf{X}; \theta)$ is an autoencoder with parameter set $\theta \in \Theta$, and $L_R(\cdot, \cdot)$ is the autoencoder’s loss function, we define the optimal parameter set for the RAE as

$$\theta^* = \operatorname{argmin}_{\theta \in \Theta} L_R \left(R(\mathcal{X}^i; \theta), \mathcal{X}^o \right), \quad (2)$$

which can be achieved through a neural network optimization process [35]. The implementation details of RAE are explained in Section 4.

3.2. Anonymization AutoEncoder

As time-windows that do not reveal sensitive inferences are supposed to pass through the RAE with minimum distortion, they may be used for other malicious purposes such as user re-identification. As a motivational example, consider

participants in a study for a new treatment who share their daily sensor data with researchers [36]. These participants may want to minimize the risk of being re-identified by those who will access their released data. Therefore, their sensor data should be released in a way that the required information for the medical study, such as patients daily activities, can be accurately inferred, while other motion patterns that facilitate user re-identification are obscured.

We define the data with the user’s identifiable information obscured as the anonymized sensor data, \mathbf{X}'' . Considering $A(\cdot)$ as a potential data transformation function and \mathbf{X} the data we want to anonymize, we define the fitness function $F(\cdot)$ as

$$F(A(\mathbf{X})) = \beta_i I(\mathbf{U}; A(\mathbf{X})) - \beta_a I(\mathbf{Y}; A(\mathbf{X})) + \beta_d d(\mathbf{X}, A(\mathbf{X})), \quad (3)$$

where the non-negative, real-valued weights β_i , β_a and β_d determine the trade-off between privacy loss and utility. As it is discussed in Section 4, the desired trade-off is established through cross validation over the training dataset.

Let the anonymization function, $\mathcal{A}(\cdot)$, that transforms \mathbf{X} into \mathbf{X}'' be

$$\mathcal{A}(\mathbf{X}) = \underset{A(\mathbf{X})}{\operatorname{argmin}} F(A(\mathbf{X})). \quad (4)$$

The threefold objective of Eq. (3) is to minimize $I(\mathbf{U}; \mathcal{A}(\mathbf{X}))$, the mutual information between the random variable that specifies the identity of current user and the anonymized data; to maximize $I(\mathbf{Y}; \mathcal{A}(\mathbf{X}))$, the mutual information between the random variable that captures the user activity and the anonymized data (*i.e.* to minimize its negative value); and, to avoid large data distortions by minimizing $d(\mathbf{X}, \mathcal{A}(\mathbf{X}))$, the distance between raw and anonymized data.

As we cannot practically search over all possible transformation functions, we consider a DNN and look for the optimal parameter set through training. To approximate the required mutual information terms, we reformulate the optimization problem in Eq. (4) as a DNN optimization problem. Let $A(\mathbf{X}; \theta)$ be a DNN, where θ is the parameter set of the DNN. The network optimizer finds the optimal parameter set θ^* by searching the space of all possible parameter sets, Θ , as:

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \beta_i I(\mathbf{U}; A(\mathbf{X}; \theta)) - \beta_a I(\mathbf{Y}; A(\mathbf{X}; \theta)) + \beta_d d(\mathbf{X}, A(\mathbf{X}; \theta)), \quad (5)$$

where $\mathcal{A}(\cdot; \theta^*)$ is the optimal data anonymizer for a general $\mathcal{A}(\cdot)$ in Eq. (4). Again, we can obtain θ^* using a stochastic optimization algorithm [37]. A key contributor to the AAE training is the following multi-objective loss function, L_A , which implements the fitness function $F(A(\mathbf{x}))$ of Eq. (4):

$$L_A = \beta_i L_i - \beta_a L_a + \beta_d L_d, \quad (6)$$

where L_a and L_d are *utility losses* that can be customized based on the target task requirements, whereas L_i is a *privacy loss* that helps the AAE remove user-specific patterns that facilitate user re-identification.

Practically, the categorical cross-entropy loss function for classification, $L_a = \mathbf{Y} \log(\hat{\mathbf{Y}})$, aims to preserve activity-specific patterns, where $\hat{\mathbf{Y}}$, the output of a softmax function, is a B -dimensional vector of probabilities for the prediction of the activity label. To tune the desired privacy-utility trade-off, the distance function that controls the amount of distortion, L_d , forces \mathbf{X}'' to be as similar as possible to the input \mathbf{X} :

$$L_d = \frac{1}{M \times W} \sum_{s=1}^M \sum_{j=1}^W (\mathbf{x}_{sj} - \mathbf{x}''_{sj})^2. \quad (7)$$

Finally, the privacy loss, L_i , the most important term of our multi-objective loss function that aims to minimize sensitive information in the data, is defined as:

$$L_i = - \left(\mathbf{U} \cdot \log(\mathbf{1}^N - \hat{\mathbf{U}}) + \log\left(1 - \max(\hat{\mathbf{U}})\right) \right), \quad (8)$$

where N is the number of users in the training set, $\mathbf{1}^N$ is the all-one column vector of length N , \mathbf{U} is the true identity label for \mathbf{X} , and $\hat{\mathbf{U}}$ is the output of the softmax function, the N -dimensional vector of probabilities learned by the classifier (*i.e.* the probability of each user label, given the input). $\mathbf{U} \cdot \log(\mathbf{1}^N - \hat{\mathbf{U}})$ is dot product of row vectors.

The goal of training AAE is to minimize the *privacy loss* by minimizing the amount of information leakage from \mathbf{U} to \mathbf{X}'' . Hence, we use adversarial training to approximate the mutual information by estimating the posterior distribution of the sensitive data given the released data [38].

4. Evaluation

To evaluate the RAE, we use four benchmark datasets of gesture and activity recognition including at least 10 different labels: Opportunity [39], Skoda [40], Hand-Gesture [41], and Utwente [42]. To evaluate the AAE, we need a dataset containing several users to show how we can hide users' gender or identity. Therefore, we use MotionSense [43] that contains the collected data of 24 users in a range of gender, age, and height who performed 6 activities. We also evaluate the compound architecture (RAE+AAE) on a case study using the MotionSense dataset.

Opportunity [39] is composed of the collected data of 4 users and there are 18 gestures classes. Each record in this dataset comprises 113 sensory readings from various types of body-worn sensors like accelerometer, gyroscope, magnetometer, and skin temperature.

Skoda [40] is collected by an assembly-line worker in a car production company wearing 19 accelerometer sensors on his right and left arm and performing a set of pre-specified experiments.

Hand-Gesture [41] includes data from accelerometer and gyroscope sensors attached to the upper and lower arm. There are two users performing 12 classes of hand movements. Each record in this dataset has 15 real-valued sensor readings.

Utwent [42] includes the data of 6 participants performing several activities, including potentially sensitive smoking activity, while wearing a smart-phone on their wrist. Accelerometer, gyroscope, and magnetometer data are collected. The whole dataset is publicly available in a single file with activity labels only.

MotionSense [43] is collected with a smart-phone kept in the users’ front pocket. A total of 24 users performed 6 activities in 15 trials in the same environment and conditions. It includes acceleration, rotation, gravity, and attitude data. Each record in this dataset includes 12 real-valued sensor readings.

Table 3 summarizes the gesture/activity classes of the five datasets. For Opportunity, we use four trials as the training data, and consider the last trial as the testing data. For other datasets, we consider 80% of the data as the training set and the rest as the testing set. The *null* class in the gesture datasets refers to data that cannot be mapped to a known behavior. All the gesture datasets are resampled to 30Hz sampling rate.

4.1. Replacement

Let the B classes of inference, $I = \{I_1, \dots, I_i, \dots, I_j, \dots, I_B\}$, be divided into three categories: (i) *Required*, $\mathcal{R} = \{I_1, \dots, I_i\}$, (ii) *Sensitive*, $\mathcal{S} = \{I_{i+1}, \dots, I_j\}$, and (iii) *Neutral*, $\mathcal{N} = \{I_{j+1}, \dots, I_B\}$. Considering a target app and its potential users, we assume \mathcal{S} is the set of inferences that users wish to keep private. We assume these are sufficiently *sensitive* that the user would wish to prevent the app from making any inferences within this set. Moreover, \mathcal{R} is the set of *required* inferences that users gain utility from if the app can accurately infer them. Finally, \mathcal{N} is the set of neutral inferences that are not sensitive to users that these inferences can be made by the app and it is also not useful for gaining utility. We assume these lists are available to the RAE for its training.

4.1.A. Gesture Datasets

Here, we implemented RAE with the following settings. Seven fully-connected layers with size (number of neurons) $inp = (M \times W), \frac{inp}{2}, \frac{inp}{8}, \frac{inp}{16}, \frac{inp}{8}, \frac{inp}{2}$, $out = inp$, respectively (except for the Hand-Gesture dataset with a lower dimensionality that the three middle layers are $\frac{inp}{3}, \frac{inp}{4}, \frac{inp}{3}$). For all datasets, we consider 1 second time-window, $W = 30$. All the experiments are performed on 30 epochs with batch size 128. The activation function for the output layer is *linear* and for the input and all of the hidden layers is Scaled Exponential Linear Unit [44]. In our experiments, to retain the overall structure of the reconstructed data, we set L_R in Eq. (2) as the point-wise mean square error function.

To evaluate the privacy loss and utility of the RAE’s outcomes, both the raw sensor data and the transformed data are given to a DNN classifier, as an envisioned app, and $F1 - score$ are calculated in Table 4, Table 5, and Table 6. Here we use $F1 - score$ as evaluation metric because it takes both false positives

Table 3: Gesture/Activity classes and properties of each dataset used for evaluation.

#	Gesture Datasets			Activity Datasets	
	Opportunity	Skoda	HandGesture	Utwente	MotionSense
0	null	null	null	—	—
1	open door1	write notes	open window	walking	standing
2	open door2	open hood	close window	jogging	stairs-down
3	close door1	close hood	water a plant	cycling	stairs-up
4	close door2	check front door	turn book	stairs-up	walking
5	open fridge	open left f door	drink a bottle	stairs-down	jogging
6	close fridge	close left f door	cut w/ knife	sitting	
7	open washer	close left doors	chop w/ knife	standing	—
8	close washer	check trunk	stir in a bowl	typing	—
9	open drawer1	open/close trunk	forehand	writing	—
10	close drawer1	check wheels	backhand	eating	—
11	open drawer2	—	smash	smoking	—
12	close drawer2	—	—	—	—
13	open drawer3	—	—	—	—
14	close drawer3	—	—	—	—
15	clean table	—	—	—	—
16	drink cup	—	—	—	—
17	toggle switch	—	—	—	—
N	4	1	2	6	24
M	113	57	15	9	12
$S.R.$	30 Hz	30 Hz	30 Hz	50 Hz	50 Hz

and false negatives into account. For RAE, false positives (*i.e.* recognizing \mathcal{R} as \mathcal{S}) harm the utility and false negatives (*i.e.* recognizing \mathcal{S} as \mathcal{R} or \mathcal{N}) harm privacy. It should be noted that the classification accuracy metric also shows similar patterns³.

The results show that utility is preserved for non-sensitive, \mathcal{R} and \mathcal{N} , classes while recognizing sensitive ones, \mathcal{S} , is very unlikely. Moreover, Figure 3 shows that the model misclassifies all transformed sections corresponding to \mathcal{S} into the \mathcal{N} and therefore the false-positive rate on *required* inferences is very low. For instance, to see how RAE can establish a good utility privacy trade-off, consider

³ Results available at: <https://github.com/mmalekzadeh/replacement-autoencoder>

Table 4: Gesture recognition results ($F1 - score$) by a pre-trained convolutional neural network on the Skoda dataset.

#	Set of Inferences	\mathbf{X}	\mathbf{X}'	#	Set of Inferences	\mathbf{X}	\mathbf{X}'
1	$\mathcal{R} = \{4, 8, 9, 10\}$	97.9	96.3	3	$\mathcal{R} = \{1, 4, 10\}$	97.6	95.0
	$\mathcal{S} = \{1, 5, 6, 7\}$	96.2	0.0		$\mathcal{S} = \{2, 3, 8, 9\}$	98.0	0.0
	$\mathcal{N} = \{0, 2, 3\}$	94.3	93.4		$\mathcal{N} = \{0, 5, 6, 7\}$	92.3	88.2
2	$\mathcal{R} = \{2, 3, 5, 6, 7, 9\}$	96.5	93.2	4	$\mathcal{R} = \{2, 3, 5, 6, 7, 9\}$	95.8	91.1
	$\mathcal{S} = \{4, 8, 10\}$	97.9	0.0		$\mathcal{S} = \{4, 8, 10\}$	97.4	0.0
	$\mathcal{N} = \{0, 1\}$	93.9	94.8		$\mathcal{N} = \{0, 1\}$	94.3	92.4

Table 5: $F1 - score$ for the Hand-Gesture dataset.

#	Set of Inferences	\mathbf{X}	\mathbf{X}'
1	$\mathcal{R} = \{1, 2, 3, 4, 9, 10, 11\}$	94.1	90.1
	$\mathcal{S} = \{5, 6, 7, 8\}$	95.7	0.3
	$\mathcal{N} = \{0\}$	95.0	96.5
2	$\mathcal{R} = \{1, 3, 4, 5, 6, 7\}$	95.2	90.4
	$\mathcal{S} = \{2, 8, 9, 10, 11\}$	94.5	0.6
	$\mathcal{N} = \{0\}$	95.0	97.5
3	$\mathcal{R} = \{1, 3, 4, 5, 6, 7, 8\}$	97.2	93.3
	$\mathcal{S} = \{2, 9, 10, 11\}$	92.5	0.7
	$\mathcal{N} = \{0\}$	95.9	97.5
4	$\mathcal{R} = \{2, 3, 5, 6, 7, 9\}$	96.1	92.1
	$\mathcal{S} = \{4, 8, 10\}$	97.0	0.5
	$\mathcal{N} = \{0, 1\}$	95.7	97.6

Table 6: $F1 - score$ for the Opportunity dataset.

#	Set of Inferences	\mathbf{X}	\mathbf{X}'
1	$\mathcal{R} = \{9, 10, \dots, 17\}$	71.8	64.3
	$\mathcal{S} = \{1, 2, \dots, 8\}$	79.1	0.2
	$\mathcal{N} = \{0\}$	88.9	89.7
2	$\mathcal{R} = \{1, 2, \dots, 8, 15, 17\}$	76.9	75.9
	$\mathcal{S} = \{9, 10, \dots, 14\}$	71.5	1.3
	$\mathcal{N} = \{0, 16\}$	84.4	82.1
3	$\mathcal{R} = \{9, 10, \dots, 14, 16\}$	74.9	77.1
	$\mathcal{S} = \{1, 2, 3, 4, 15, 17\}$	76.2	0.9
	$\mathcal{N} = \{0, 5, 6, 7, 8\}$	85.0	81.6
4	$\mathcal{R} = \{1, 2, \dots, 8, 15, 17\}$	70.3	65.0
	$\mathcal{S} = \{9, 10, \dots, 14, 16\}$	74.9	6.3
	$\mathcal{N} = \{0, 1\}$	93.7	92.9

the results for the Skoda dataset in Table 4(#2). We see that the gesture classifier can effectively recognize \mathcal{R} gestures (*e.g.* opening and closing doors), even when the app processes the output of RAE instead of the raw data (with 93.2% accuracy). However, \mathcal{S} gestures (*e.g.* checking doors) that can be recognized with high accuracy when app processes the raw data (with 97.9% accuracy), are completely filtered out in the output of the RAE. Moreover, the corresponding confusion matrix for this experiment in Figure 3 (Left-Bottom) shows that the utility of the *required* inferences is preserved as the classifier wrongly infers all \mathcal{S} gestures as some \mathcal{N} gestures and not as one of \mathcal{R} gestures.

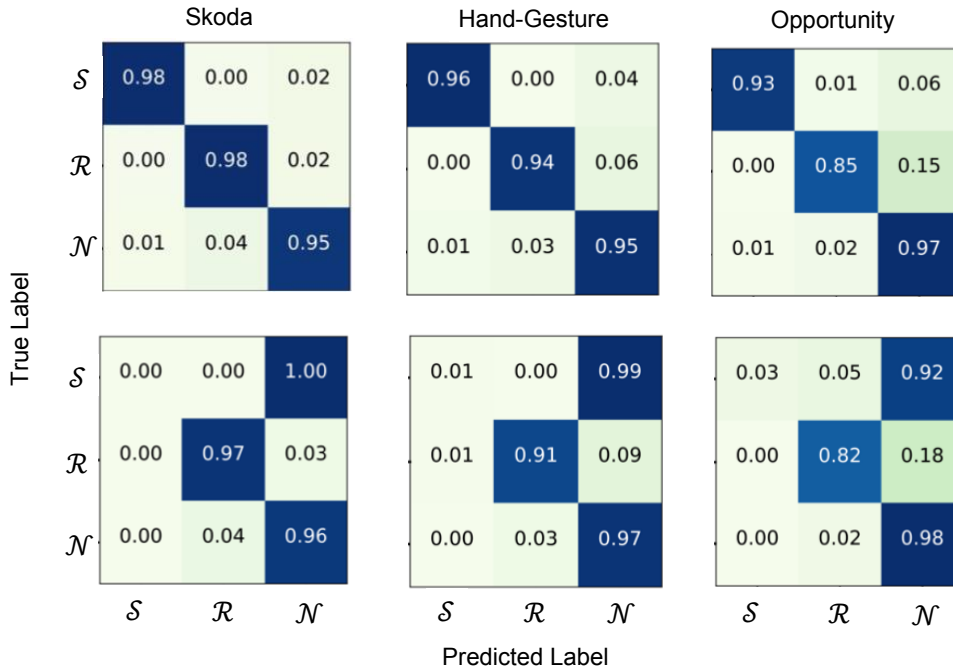


Figure 3: Confusion Matrix for (top) original time-series, and (bottom) transformed time-series by RAE. After transformation almost all the *sensitive* gestures are recognized as *neutral* ones. (Left) Results on the Skoda dataset in Table 4 (#2). (Middle) Hand-Gesture dataset in Table 5 (#1). (Right) Opportunity dataset in Table 6 (#2).

4.1.B. Utwente dataset

Let $\mathcal{S} = \{\text{typing, writing, smoking, eating}\}$ be the set of sensitive inferences, $\mathcal{N} = \{\text{sitting still, standing still}\}$ be the set of neutral inferences, and $\mathcal{R} = \{\text{walking, jogging, cycling, stairs-up, stairs-down}\}$ be the set of required inferences. Considering a 2-second time-window ($W = 100$), we trained an RAE with 6 hidden layers: 4 Convolutional-LSTM layers using *hyperbolic tangent* as the activation function with 256, 128, 64 and 64 filters respectively, followed by 2 Convolutional layers using Scaled Exponential Linear Unit [44] as the activation function with 64 and 128 layers respectively. We also put a batch-normalizer on the output of each hidden layer to reduce the training time⁴.

To evaluate the privacy-utility trade-off, we use a DNN classifier. As we see in Table 7, the average accuracy of the classifier on the raw data is more than 99%. However, when we feed the same classifier with the output of the RAE, all the \mathcal{S} activities are recognized as *sitting still*, while the accuracy for \mathcal{R} activities is almost equal to that of the raw data. We observe that for *smoking* there still is a 5% chance of recognition. Note that for some time-windows of the *smoking* the raw data are similar to those of *standing still*. This effect can be also seen

⁴More implementation details and codes for reproducing the results can be found at <https://github.com/mmalekzadeh/replacement-autoencoder>

	walking	jogging	cycling	stairs-up	stairs-down	sitting	standing	typing	writing	eating	smoking
walking	97.5 → 97.2			0.7 → 0.7	1.5 → 1.9						0.3 → 0.1
jogging		100 → 100									
cycling			100 → 100								
stairs-up	0.4 → 0.3	0.4 → 0.4	0.0 → 0.1	98.8 → 98.8			0.1 → 0.1				0.3 → 0.3
stairs-down				0.3 → 0.3	99.7 → 99.7						
sitting			0.0 → 0.3			98.6 → 96.8		1.0 → 0.0	0.1 → 0.0	0.1 → 0.0	0.1 → 2.8
standing			0.0 → 0.3				99.4 → 98.2				0.6 → 1.5
typing						0.0 → 100		100 → 0.0			
writing			0.0 → 0.7			0.0 → 99.3			99.9 → 0.0	0.1 → 0.0	
eating			0.0 → 0.5			0.1 → 99.4	0.3 → 0.0			99.6 → 0.0	0.1 → 0.1
smoking			0.0 → 0.1			0.0 → 94.9				2.3 → 0.0	97.5 → 5.0

Table 7: Confusion Matrix of the results on the test data for Utwente dataset. Rows show the true labels and columns show the predicted labels. In each cell, the left part shows the accuracy on the raw data, and the right part shows the accuracy after transformation. For brevity, all the values are rounded to one decimal point. Empty cells show 0.0 → 0.0.

in Table 7 (column *smoking*). We assume this is a labeling error when the data curator labels intervals between cigarette drags as smoking behavior while user is standing.

4.2. Anonymization

To evaluate the AAE as a data anonymizer, we measure the extent to which the accuracy of activity recognition suffers from anonymization compared to accessing the raw data. We compare the trade-off between recognizing users’ activity versus their identity, and compare with baseline methods for coarse-grain time-series data (*resampling* and *singular spectrum analysis*) and with the method in [13] that only considers the latent representation by the Encoder model (see Figure 4), without taking \mathbf{X}'' into account.

We use resampling by Fast Fourier Transform (FFT), which is desirable for periodic time-series, and this is typical with mobile sensor data for activity recognition. Singular Spectrum Analysis (SSA) [45] is a model-free technique that decomposes time-series into trend, periodic, and structureless (or noise) components using singular value decomposition (SVD). In our case, we decompose $\mathbf{X} = \{\mathbb{X}^1, \mathbb{X}^2, \dots, \mathbb{X}^D\}$ such that the \mathbb{X}^i and \mathbb{X}^{i+1} are arranged in descending order according to their corresponding singular value and the original time-series can be recovered as: $\mathbf{X} = \sum_{i=1}^D \mathbb{X}^i$. Thus, we test the idea of *incremental reconstruction* by SSA as a base-line transformation method.

We consider two methods of dividing the dataset into training and test sets, namely *Subject* and *Trial*. For *Subject*, we put all data of 4 of the users in the dataset, 2 females and 2 males, as testing data and the remaining 20 users as training. Hence, after training the AAE, we evaluate the model on a dataset of new unseen users. For *Trial*, we put one trial data of each user as testing data and the remaining trials of that user’s data as training. For example, where we have three walking trials for every user, we consider one trial as testing and the other two as training. In both cases, we put 20% of the training data for validation during the training phase. We repeat each experiment 5 times and report the

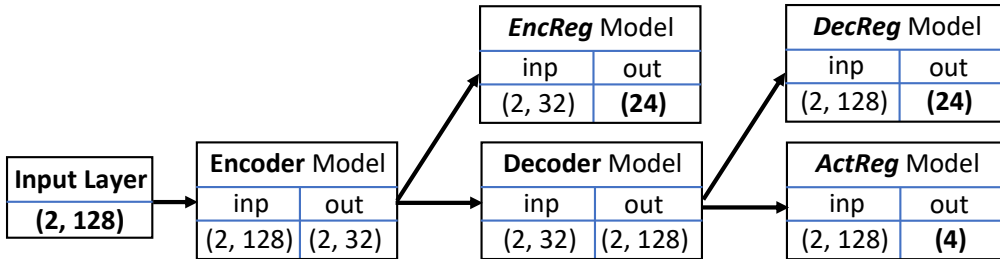


Figure 4: Details of the training the AAE (Encoder and Decoder) for a dataset with 24 users and 4 activities. KEY – *EncReg* and *DecReg*: users’ identity recognizers that monitors the output of the Encoder and Decoder, respectively, to reduce the privacy loss; *ActReg*: a users’ activity recognizer that monitors the output of the Decoder to increase the utility.

mean and the standard deviation. For all the experiments we use the magnitude value for both gyroscope and accelerometer.

To simplify the process of encoding data into a lower-dimensional representation and then decoding it to the original dimension with convolutional filters, we set W to be a power of 2. The larger W , the lower the possibility of taking advantage of the correlation among the successive windows by adversaries [38]. But larger window sizes increase the delay for real-time apps. We set $W = 128$ (*i.e.* 2.56 seconds).

For all the regularizers, *EncReg*, *DecReg*, and *ActReg* (see Figure 4), we use 2D convolutional neural networks. To prevent overfitting, we add a Dropout [46] layer after each convolution layer. We also use L2 regularization to penalize large weights. We train the classifier on the original and the anonymized training dataset, and then use it for inference on the test data. We use the Subject setting, thus the test data includes data of new unseen users.

To measure the utility, we train an activity recognition classifier on both the raw data and the output of each transformation method: *Resampling*, *SSA*, [13], and our *AAE*. Then, we use the trained model for inference on the corresponding testing data. Here we use the *Subject* setting, thus the testing data include data of new unseen users. The second row of Table 8 (ACT) shows that the average accuracy for activity recognition for both Raw and AAE data is around 92%. Compared to other methods that decrease the utility of the data, we can preserve the utility and even slightly improve it, on average, as the AAE shapes data such that an activity recognition classifier can learn better from the transformed data than from the raw data.

To measure the privacy loss, we assume that an adversary has access to the training dataset and we measure the ability of a pre-trained deep classifier on users raw data in inferring the identity of the users when it receives the transformed data. We train a classifier in the *Trial* setting over raw data and then feed it different types of transformed data. The third row of Table 8 (ID) shows that downsampling data from 50Hz to 5Hz reveals more information than using the AAE output in the original frequency. These results show that the AAE can

Table 8: Trade-off between utility (activity recognition) and privacy (protecting identity). The forth row shows the K-NN rank between 24 users (the lower the better). Key – *ACT*: activity recognition, *ID*: identity recognition, *ACC*: accuracy, *F1*: *F1 – score*, *DTW*: Dynamic Time Warping as similarity measure, *SSA*: Singular Spectrum Analysis, *AAE*: Our method.

<i>Experiment</i>	<i>Measure</i>	<i>Raw Data</i>	<i>Resampling</i>			<i>SSA</i>		[13]	<i>AAE</i>
		<i>50Hz</i>	<i>10Hz</i>	<i>5Hz</i>	<i>1+2</i>	<i>1</i>	<i>50Hz</i>	<i>50Hz</i>	
<i>ACT</i>	mean F1	92.5	91.1	88.0	88.6	87.4	91.5	92.9	
	xiance F1	2.1	0.6	1.8	0.9	0.9	0.9	0.37	
<i>ID</i>	mean ACC	96.2	31.1	13.5	34.1	16.1	15.9	7.0	
	mean F1	95.9	25.6	8.9	28.6	12.6	11.2	1.8	
<i>DTW</i>	mean Rank	0	7.2	9.3	6.8	9.5	10.7	6.6	
	variance Rank	0	5.7	5.8	5.6	5.4	5.5	4.7	

effectively obscure user-identifiable information so that even a model that has had access to the original data of the users cannot distinguish them after applying the transformation.

Finally, to evaluate the privacy loss and efficiency of the anonymization with an unsupervised mechanism, we implement the k -Nearest Neighbors (k -NN) with Dynamic Time Warping (DTW) [47]. Using DTW, we measure the similarity between the transformed data of a target user k and the raw data of each user l , \mathbf{X}^l , for all $l \in \{1, \dots, k, \dots, N\}$. Then we use this similarity measure to find the k nearest neighbors of user l and check their rank. The last row of Table 8 (DTW) shows that it is very difficult to find similarities between the transformed and raw data of the users as the performance of the AAE is very similar to the baseline methods and the constraint in Eq. (5) maintain the data as similar as possible to the original data. This result shows that the utility-privacy trade-off of AAE is preferable to that of the other methods.

4.3. Compound Architecture

Here, we evaluate a setting where anonymization with the AEE follows replacement using RAE. Considering MotionSense dataset, we want an app to be unable to infer gender or jogging activity from motion data. Let $\mathcal{S}=\{\text{jogging}\}$ be the *sensitive* activity to be replaced with $\mathcal{N}=\{\text{standing still}\}$ as the *neutral* activity. We also consider $\mathcal{R}=\{\text{walking, stairs-down, stairs-up}\}$ as the *required* inferences. Let the time-window be 2.56 seconds ($W = 128$ samples) and $M = 2$, *i.e.* we consider the magnitude of rotation and acceleration of the device.

First, we train two convolutional neural networks as activity and gender classifiers on the original training dataset. Second, RAE is trained to replace the jogging time-windows while keeping the *required* time-windows unmodified in the RAE’s output, \mathbf{X}' . Third, we use the RAE’s output, \mathbf{X}' , as the AAE’s input and train the AEE to reduce the likelihood of the user’s gender being inferred from

Table 9: True-positive rate for each activity and gender classification accuracy (%) using a convolutional neural network for each stage of the compound model on MotionSense [43] dataset.

<i>Inference</i>	\mathbf{X} : <i>Original</i>	\mathbf{X}' : <i>Replacement</i>	\mathbf{X}'' : <i>Anonymization</i>		
			$\beta_i = \beta_a = \beta_d$	$\beta_i = \frac{1}{2}\beta_a = \beta_d$	
	stairs-down	98.0	93.9	98.5	96.3
\mathcal{R}	stairs-up	96.4	97.8	92.3	96.3
	walking	99.7	94.8	89.4	94.8
\mathcal{S}	jogging	99.3	1.4 (92 as \mathcal{N})	.2 (92 as \mathcal{N})	.1 (84 as \mathcal{N})
\mathcal{N}	standing	99.9	99.9	100	99.9
Gender		98.9	97.1	45.0	39.0

the ultimate data that is shared with the app, \mathbf{X}'' . Finally, after training both autoencoders, we feed the testing dataset into the compound model.

Table 9 shows the activity and gender classification results at each processing stage. While \mathbf{X} is highly informative for all inferences, after replacement *jogging* intervals are not inferred in \mathbf{X}' and they are classified as *standing*. However, *gender* can still be inferred from \mathbf{X}' . Inferring gender from \mathbf{X}'' (*i.e.* after anonymization) reaches the desired level of random guess while the inference of \mathcal{R} is maintained close to the original accuracy. Importantly, the proposed framework allows us to give different weights on preserving the activity and hiding gender: the last column of Table 9 shows that a better accuracy can be obtained if we increase the risk of leaking more sensitive information. Notice that, the random guess is 50% accurate. Thus, the privacy loss is larger when we have 39% accuracy for gender classifier than 45%.

5. Discussion

While we believe the proposed mechanisms establish effective utility-privacy trade-offs for sensor data transformations, here we discuss directions of this work that need more explorations.

First, in the available datasets, the activities/gestures that are categorized into *sensitive*, *required*, and *neutral* are independent of each other and at each time-window only one of them is happening. However, in the real-world situations there might be correlations among different activities that affect the provided privacy guarantees for some sensitive inferences. Similarly, correlations among consecutive time-windows of a specific activity may incrementally reveal information that facilitate user re-identification. To assess this, we would need access to multi-labelled data collected over a much longer time period as well as a large number of demographically different users.

Second, to show that the proposed mechanisms can generalize, we performed evaluations on several datasets collected from different type of sensors located

in different part of users' body. Current public datasets of mobile and wearable sensor data do not simultaneously satisfy the requirements of abundance and variety of activities and users. To reduce the risk of overfitting, we performed our experiments on DNNs with small number of layers and small number of neurons in each layer. With larger datasets, one can increase the learning capacity of the RAE and AAE by adding more layers to the neural network or investigate various DNN architectures.

Third, we have assumed the existence of a publicly available dataset to train the RAE and the AAE. when such public dataset is not available, one option is to use privacy-preserving model training without collecting personal data [48], or training the required model through a federated learning [49].

Finally, we aim to investigate a privacy-preserving mechanism that transforms sensitive patterns into a mixture of neutral activities rather than only one of them. Moreover, we aim to look for, or to collect, larger datasets to conduct experiments on additional tasks, to derive statistical bounds for the amount of privacy achieved, and to measure the cost of running the proposed local transformations on user devices.

6. Conclusion

In this paper we showed how to achieve a trade-off between privacy and utility for sensor data release with an appropriate learning process. In particular, we presented new ways to train deep autoencoders for continuous data transformations to prevent a honest-but-curious app from discovering users' sensitive information. Our model is general and can be applied to unseen data of new users, without need for re-training. Experiments conducted on various types of real-world sensor data showed that our transformation mechanism eliminates the possibility of making sensitive inferences and obscures user-specific motion patterns that enable user re-identification, introducing a small utility loss for activity and gesture recognition tasks.

Acknowledgment

The work was supported by the Life Sciences Initiative at Queen Mary University of London and a Microsoft Azure for Research Award (CRM:0740917). Andrea Cavallaro wishes to thank the Alan Turing Institute (EP/N510129/1), which is funded by the EPSRC, for its support through the project PRIMULA. Hamed Haddadi was partially supported by the EPSRC Databox grant (EP/N028260/1).

References

- [1] K. Katevas, H. Haddadi, L. Tokarchuk, R. G. Clegg, Walking in sync: Two is company, three's a crowd, in: Proceedings of the 2nd Workshop on Workshop on Physical Analytics, WPA '15, Florence, Italy, ACM, 2015, pp. 25–29.

- [2] K. Hänsel, K. Katevas, G. Orgs, D. C. Richardson, A. Alomainy, H. Haddadi, The potential of wearable technology for monitoring social interactions based on interpersonal synchrony, in: Proceedings of the 4th ACM Workshop on Wearable Systems and Applications, ACM, 2018, pp. 45–47.
- [3] M. Irfan, L. Tokarchuk, L. Marcenaro, C. Regazzoni, Anomaly detection in crowds using multi sensory information, in: 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), IEEE, 2018, pp. 1–6.
- [4] P. M. Scholl, K. Van Laerhoven, A feasibility study of wrist-worn accelerometer based detection of smoking habits, in: 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IEEE, 2012, pp. 886–891.
- [5] Q. Riaz, A. Vögele, B. Krüger, A. Weber, One small step for a man: Estimation of gender, age and height from recordings of one step by a single inertial sensor, *Sensors* 15 (12) (2015) 31999–32019.
- [6] N. Neverova, C. Wolf, G. Lacey, L. Fridman, D. Chandra, B. Barbello, G. Taylor, Learning human identity from motion patterns, *IEEE Access* 4 (2016) 1810–1820.
- [7] A. F. Westin, Privacy and freedom, *Washington and Lee Law Review* 25 (1) (1968) 166.
- [8] A. Ghosh, T. Roughgarden, M. Sundararajan, Universally utility-maximizing privacy mechanisms, *SIAM Journal on Computing* 41 (6) (2012) 1673–1693.
- [9] C. Dwork, A. Roth, The algorithmic foundations of differential privacy, *Foundations and Trends in Theoretical Computer Science* 9 (3–4) (2014) 211–407.
- [10] A. Bittau, U. Erlingsson, P. Maniatis, I. Mironov, A. Raghunathan, D. Lie, M. Rudominer, U. Kode, J. Tinnes, B. Seefeld, Prochlo: Strong privacy for analytics in the crowd, in: Proceedings of the 26th Symposium on Operating Systems Principles, ACM, 2017, pp. 441–459.
- [11] S. Menasria, J. Wang, M. Lu, The purpose driven privacy preservation for accelerometer-based activity recognition, *World Wide Web* 21 (6) (2018) 1773–1785.
- [12] C. Huang, P. Kairouz, X. Chen, L. Sankar, R. Rajagopal, Context-aware generative adversarial privacy, *Entropy* 19 (12) (2017) 656.
- [13] H. Edwards, A. J. Storkey, Censoring representations with an adversary, in: 4th International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, May 2–4, 2016.
- [14] G. Miklau, D. Suciu, A formal analysis of information disclosure in data exchange, *Journal of Computer and System Sciences* 73 (3) (2007) 507–534.
- [15] F. du Pin Calmon, N. Fawaz, Privacy against statistical inference, in: 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2012, pp. 1401–1408.
- [16] A. Ghosh, R. Kleinberg, Inferential Privacy Guarantees for Differentially Private Mechanisms, in: 8th Innovations in Theoretical Computer Science Conference (ITCS), Dagstuhl, Germany, 2017, pp. 9:1–9:3.
- [17] Y. Amar, H. Haddadi, R. Mortier, An information-theoretic approach to time-series data privacy, in: Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems, ACM, 2018, p. 3.
- [18] H. Wang, Z. Xu, CTS-DP: Publishing correlated time-series data via differential privacy, *Knowledge-Based Systems* 122 (2017) 167–179.
- [19] T. Zhu, P. Xiong, G. Li, W. Zhou, Correlated differential privacy: hiding information in non-iid data set, *IEEE Transactions on Information Forensics and Security* 10 (2) (2015) 229–242.
- [20] P. K. Diederik, M. Welling, et al., Auto-encoding variational bayes, in: Proceedings of the International Conference on Learning Representations (ICLR), 2014.
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2014, pp. 2672–2680.
- [22] B. K. Beaulieu-Jones, Z. S. Wu, C. Williams, R. Lee, S. P. Bhavnani, J. B. Byrd, C. S. Greene, Privacy-preserving generative deep neural networks support clinical data sharing, *Circulation: Cardiovascular Quality and Outcomes* 12 (7) (2019) e005122.

- [23] G. Acs, L. Melis, C. Castelluccia, E. De Cristofaro, Differentially private mixture of generative neural networks, *IEEE Transactions on Knowledge and Data Engineering* 31 (6) (2018) 1109–1121.
- [24] F. Laforet, E. Buchmann, K. Böhm, Individual privacy constraints on time-series data, *Information Systems* 54 (2015) 74–91.
- [25] C. Esteban, S. L. Hyland, G. Rättsch, Real-valued (medical) time series generation with recurrent conditional gans, *arXiv preprint 1706.02633*, 2017.
- [26] M. Götz, S. Nath, J. Gehrke, MaskIt: Privately releasing user context streams for personalized mobile applications, in: *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD '12*, ACM, 2012, pp. 289–300.
- [27] N. Saleheen, S. Chakraborty, N. Ali, M. M. Rahman, S. M. Hossain, R. Bari, E. Buder, M. Srivastava, S. Kumar, msieve: differential behavioral privacy in time series of mobile sensor data, in: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 706–717.
- [28] S. A. Osia, A. Taheri, A. S. Shamsabadi, M. Katevas, H. Haddadi, H. R. Rabiee, Deep private-feature extraction, *IEEE Transactions on Knowledge and Data Engineering*.
- [29] N. Raval, A. Machanavajjhala, J. Pan, Olympus: Sensor privacy through utility aware obfuscation, *Proceedings on Privacy Enhancing Technologies* 1 (2019) 21.
- [30] A. S. Shamsabadi, H. Haddadi, A. Cavallaro, Distributed one-class learning, in: *25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 4123–4127.
- [31] M. Lu, Y. Guo, D. Meng, C. Li, Y. Zhao, An information-aware privacy-preserving accelerometer data sharing, in: *International conference of pioneering computer scientists, engineers and educators*, Springer, Singapore, 2017, pp. 425–432.
- [32] F. Xiao, M. Lu, Y. Zhao, S. Menasria, D. Meng, S. Xie, J. Li, C. Li, An information-aware visualization for privacy-preserving accelerometer data sharing, *Human-centric Computing and Information Sciences* 8 (1) (2018) 13.
- [33] I. Psychoula, E. Merdivan, D. Singh, L. Chen, F. Chen, S. Hanke, J. Kropf, A. Holzinger, M. Geist, A deep learning approach for privacy preservation in assisted living (2018) 710–715.
- [34] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, *IEEE transactions on pattern analysis and machine intelligence* 35 (8) (2013) 1798–1828.
- [35] M. Malekzadeh, R. G. Clegg, H. Haddadi, Replacement autoencoder: A privacy-preserving algorithm for sensory data analysis, in: *2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI)*, IEEE, 2018, pp. 165–176.
- [36] D. Rodriguez-Martin, C. Perez-Lopez, A. Sama, A. Catala, J. Moreno Arostegui, J. Cabestany, B. Mestre, S. Alcaine, A. Prats, M. Cruz Crespo, A. Bayes, A waist-worn inertial measurement unit for long-term monitoring of parkinsons disease patients, *Sensors* 17 (4) (2017) 827.
- [37] D. P. Kingma, B. J. Adam, A method for stochastic optimization. *arxiv e-prints*, 2014, in: *The 3rd International Conference for Learning Representations*, San Diego, 2015.
- [38] M. Malekzadeh, R. G. Clegg, A. Cavallaro, H. Haddadi, Mobile sensor data anonymization, in: *Proceedings of the International Conference on Internet of Things Design and Implementation (IoTDI)*, ACM, 2019, pp. 49–58.
- [39] R. Chavarriaga, H. Sagha, A. Calatroni, S. T. Digumarti, G. Tröster, J. d. R. Millán, D. Roggen, The opportunity challenge: A benchmark database for on-body sensor-based activity recognition, *Pattern Recognition Letters* 34 (15) (2013) 2033–2042.
- [40] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, G. Troster, Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection, *Lecture Notes in Computer Science* 4913 (2008) 17.
- [41] A. Bulling, U. Blanke, B. Schiele, A tutorial on human activity recognition using body-worn inertial sensors, *ACM Computing Surveys* 46 (3) (2014) 33:1–33:33.
- [42] M. Shoaib, S. Bosch, O. Incel, H. Scholten, P. Havinga, Complex human activity recognition using smartphone and wrist-worn motion sensors, *Sensors* 16 (4) (2016) 426.

- [43] M. Malekzadeh, R. G. Clegg, A. Cavallaro, H. Haddadi, Protecting sensory data against sensitive inferences, in: Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems, W-P2DS'18, ACM, 2018, pp. 2:1–2:6.
- [44] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, Self-normalizing neural networks, in: Advances in Neural Information Processing Systems, 2017, pp. 971–980.
- [45] D. S. Broomhead, G. P. King, Extracting qualitative dynamics from experimental data, *Physica D: Nonlinear Phenomena* 20 (2-3) (1986) 217–236.
- [46] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research* 15 (1) (2014) 1929–1958.
- [47] S. Salvador, P. Chan, Toward accurate dynamic time warping in linear time and space, *Intelligent Data Analysis* 11 (5) (2007) 561–580.
- [48] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, L. Zhang, Deep learning with differential privacy, in: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, ACM, 2016, pp. 308–318.
- [49] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konecny, S. Mazzocchi, H. B. McMahan, T. Van Overveldt, D. Petrou, D. Ramage, J. Roselander, Towards federated learning at scale: System design, in: Proceedings of the 2nd SysML Conference, Palo Alto, CA, USA, 2019.