Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-13185-4

Differentially private human activity recognition for smartphone users

Avishek Garain[1] · Rudrajit Dawn[1] · Saswat Singh[1] · Chandreyee Chowdhury[1]

Received: 16 April 2021 / Revised: 4 March 2022 / Accepted: 19 April 2022

Abstract
User privacy is an important concern that should be handled in data intensive applications. Interestingly, differential privacy is a privacy model that can be applied to such datasets. This model is advantageous as it does not make any strong assumption about the adversary. In this work, we have introduced the notion of differential privacy in the domain of Human Activity Recognition (HAR). Real life accelerometer data has been collected from different smartphone configurations that were carried by the users in different manner according to their convenience. Our contribution in this work is to propose a privacy preserving HAR framework incorporating algorithms to preserve the differential privacy of the user data. The algorithm exploits the scalar and the vector parts of the accelerometer readings and applies privacy preserving mechanisms on it. A Deep Multi Layer Perceptron (DMLP) framework has been utilized for activity classification. We have achieved comparatively similar results with an enhanced surplus of achievement of privacy in terms of data and are so far the first of its kind in the aforementioned domain of HAR based on smartphone sensing data. The proposed framework is implemented both on collected real life dataset capturing different smartphone configurations and usage behavior and benchmark datasets.

Keywords HAR · Noise · Privacy · Differential privacy · Deep MLP · Smartphones

✉ Chandreyee Chowdhury
chandreyee.chowdhury@gmail.com

Avishek Garain
avishekgarain@gmail.com

Rudrajit Dawn
rudrajit.dawn@gmail.com

Saswat Singh
singhsaswat02@gmail.com

[1] Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

Published online: 13 May 2022

1 Introduction

Differential privacy has been designed to protect the privacy between neighboring datasets which differ in only one element [6]. It means that the adversary could not distinguish whether one of the elements changes based on the releasing result. The major strength of this privacy model is that it does not make any strong assumptions about the adversary and hence, can be applied to various forms of datasets.

In recent times, there is an emergence of assorted networks and applications involving Human Activity Recognition (HAR). Along with the emergence of sensor-rich smartphones and their connectivity to cloud servers over the Internet, many interesting applications are emerging to evolve based on HAR [16, 25]. The availability of inertial sensors and GPS in widely used smartphone models presents unprecedented opportunities for the collection of data to study human behavior both individual and social and hence, develop societal applications for a better life. It has applications ranging from smart healthcare to fitness tracking to entertainment and human-computer interactions as well. It is a step toward proactive healthcare where the health conditions of older adults could be monitored at home and hence, the cost of caregiving and hospitalization can be immensely reduced. Machine learning and deep learning techniques have been the driving force behind these sets of applications whose major emphasis is on the analysis of user data collected from smart devices, such as, smartphones, smartwatches, and so on. Thus, it has led to the generation of tremendous amounts of data containing users' behavioral information as well. However, some attributes of the data may be sensitive and should not be divulged without users' consent. This in turn leads to the demand for inclusion of differential privacy in these domains that too at an alarming rate. Privacy is one of the major concerns that need to be addressed for the wide adoption of such emerging data driven applications.

A simple technique is ID removal which has proven its vulnerability to de-anonymization attacks [8]. K-anonymity-based techniques, as in [23], are suitable for anonymization of relational data, however, they are only designed to anonymize some specific structural semantics, and can often be overcome by other structural semantics. Fortunately, differential privacy based techniques, which theoretically provide a strong privacy guarantee, have come to the rescue to solve the vulnerability. The usage of methods like the dK-2 series feature extraction model or the Hierarchical Random Graph (HRG) model have shown promising results [3]. These algorithms have achieved global -differential privacy over the entire dataset with different graph abstraction models [28]. But network data is very sensitive to the changes in network structure. Although these global differential privacy techniques are rich in preserving privacy, the regenerated graph lacks enough utility for network data analysis. How to balance the anonymization technique's privacy level with its negative utility impact is always a question for privacy-preserving data publishing. Anonymity is not always a complete solution for privacy. Machine learning models often reveal sensitive data about the users as well [21]. Thus, the schemes related to differential privacy abandon the attempt to exploit a user's information obtained from various fitness devices and extracting valuable information from the same. Then they can preserve more data utility while keeping the privacy level higher than the pure local-differential privacy criteria as in [9]. Though a lot of work is going on for smartphone sensing applications and the need for privacy is also highlighted in a few works [27], fewer attempts [2] are being made to preserve the privacy of smartphone sensing data, especially for HAR. This has motivated us to explore the notion of differential privacy in this context.

In our work, we have introduced these privacy preserving principles to the HAR domain. We have made use of a Deep Multi-Layer Perceptron (DMLP) classifier for the purpose of activity classification. Noise generated using various parameters in the data is inserted into the dataset, such that the user data remains differentially private from the owners of the application that is responsible for generating the data. Algorithms are designed for the data pre-processing and feature extraction purpose. A detailed analysis of numerous types of data perturbation mechanisms that might be applied to the data without penalizing the classification accuracy is investigated. We have also tried to figure out if confidential user information can be extracted from the data employing other attributes present in the data and analyzed the same in detail.

The rest of the paper has been organized as follows. State of the art literature is studied in Section 2 followed by a depiction of differential privacy in Section 3. Section 4 describes the data, on which, the task was performed. The methodology followed is described in Section 5. This is followed by the results and concluding remarks in Sections 6 and 7, respectively.

2 Literature survey

More than 3 billion mobile devices are in use nowadays [1], with multiple sensors (e.g., accelerometer, gyroscope, and GPS) that can capture detailed and objective measurements on various aspects of our lives, including physical activity, health status, location both indoor and outdoor. Privacy is an important factor for such data driven applications as deep learning analysis techniques may divulge sensitive information based on these behavioral datasets. Imprecise data annotation in sensor based services is often viewed as a primitive step for privacy in literature particularly in the context of smartphone sensing applications. For instance, in location sensitive applications, approaches are found that hinder the precise location of the user. Without any protocol on preserving the user data, the service providers can gather and record the precise location data of the user and trade them to 3rd parties. Many privacy mechanisms have been proposed for users.

In [12], Liu et al. investigated the predictability of privacy on mobile crowdsensing data, which they envisioned to have capabilities to measure the privacy protections along with giving access to application users to forecast the loss in utility at an equivalent time. They proposed the Salus algorithm to ensure that private data is protected against data reconstruction attacks. Furthermore, using Salus they provide accurate utility forecasts for crowdsensing applications to predict the utility. The assessments of risk are often generally applied to sensors of different types on the mobile platform, and therefore, the utility prediction can also be used in applications that use data aggregators like average, histogram, and classifiers to provide support. In the work of Deloc [18], a delegation-based privacy-preserving mechanism for location-based services is presented by Sahmoune et al. They have devised a mechanism to protect the location privacy of the user without altering the coordinates.

Many algorithms provide a partial disclosure of the user's location. Shun et al. in [20] have worked on the differential privacy that deals with dynamic location obfuscation with personalized error bounds. They have shown that PIVE fails to offer differential privacy on adaptive protection of location set as proclaimed. According to them, for location privacy,

[1] https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/

Geo-indistinguishability and expected inference error are two complementary notions. They have demonstrated that the intersection of different location set with one another could be due to the defined search algorithm and the different locations that are in the same protection location set could have different protection diameters.

Differential Privacy also finds its application in the Cyber-physical system (CPS), an evolving technology. Privacy is a primary security requirement in CPS and can cause serious damages if unresolved. Much work is completed within the area of privacy preservation in CPS.

Agarwal et al. in [1], have worked on a generic mechanism for privacy preservation in CPS. Their proposed study has aimed to integrate separate privacy protection mechanisms in different levels of the CPS architecture, addressing different kinds of privacy as information contents, locations, identities, dates and times, addresses, etc., within a common structure. However, it is more challenging to balance the privacy requirement and system performance on complex CPSs. The reason is that a complex CPS has a dynamic space-time coupling. So, to further optimize space-time, it will lead to increase in system cost.

In [24], Tran et al. have constructed an efficient approach for privacy preserving models based on secure multi-party computation using deep learning. It follows the federated learning approach utilizing the concept of decentralized training and thus, ensuring the privacy of local user data. They first proposed a so-called Efficient Secure Sum Protocol (ESSP) that provides a different group of parties to calculate the sum of private data inputs. ESSP works with integer numbers as well as with floating-point numbers with no conversion. A proposed Secure Model Sharing Protocol permits groups of parties to securely train and share the local models which aggregate into a worldwide model. They conducted theoretical calculations of privacy and communication cost as well as empirical experiments on balance class image datasets and an unbalance class text dataset. They even compared their approach theoretically with perturbation mechanisms employed in differential privacy based models.

The authors Kasiviswanathan et al. (2013) [10] have developed algorithms that provide accurate analysis of realistic networks for the private analysis of network data, as well as satisfying privacy guarantees. They have presented several techniques for designing nodes and for differentially private algorithms, that is, algorithms whose output distribution does not reflect significant changes when nodes and edges are added to a graph. They have developed a methodology for analyzing the accuracy of developed algorithms on realistic networks. The projection operators have been tailored to specific statistics with low sensitivity and preserve information about the first statistic. Additionally, they have derived a generic and more efficient reduction that allows them to use any differentially private algorithm for bounded-degree graphs to an arbitrary graph. This reduction is predicated on analyzing the sensitivity of the simple truncation that discards nodes of a higher degree.

If we consider the domain of HAR, some privacy-aware works have already been developed and are already in use.

In the work by Ryoo et al. [15], the authors have presented a quite rudimentary approach for addressing objectives that are quite contradicting, namely human activity recognition while only using extreme low-resolution anonymized videos. They have introduced the paradigm of inverse super resolution (ISR), which involves learning the optimal set of image transformations for generating multiple low-resolution (LR) training videos from a single video. Their ISR learns various types of sub-pixel transformations which are optimized for the main purpose of classification of activity; thus allowing the classifier for taking the best advantage of existing high-resolution videos employing creating multiple LR training videos created for the problem. The authors have experimentally confirmed that the

paradigm of inverse super resolution has been able to prove itself beneficial for activity recognition from extreme low-resolution videos.

The authors Fredrikson et al. [5] in their work have developed a comparatively newer type of model inversion attack that tries to exploit confidence values revealed along with predictions. Their new attacks are applicable in various types of settings. Out of which, they have explored two settings in depth namely, decision trees for lifestyle surveys used on systems powered by machine learning as a service and neural networks for facial recognition. For both the use cases, the confidence values are revealed to those with the ability to make prediction queries to models. The authors have experimentally shown attacks that can predict if a respondent in a lifestyle survey admitted to cheating on their significant other. In case of facial recognition, they have shown the process for recovering recognizable images of people's faces given only their name and access to the machine learning model. They have also done experimental exploration of countermeasures, thus, experimenting and modeling a privacy-aware decision tree classifier which is a variant of Classification And Regression Trees (CART) learning.

The authors Samarah et al. [19] in their work have proposed an architecture that does efficient recognition of human activities in smart homes based on the method of Spatio-temporal mining. They have also presented a method for enhancing the privacy of the collected human sensed activities by making use of a slightly modified version of the micro-aggregation based approach. The authors have performed an extensive validation of their framework on benchmark data sets which in turn yielded highly promising results in terms of accuracy and privacy-utility trade off.

In the work by Zheng et al. [29], the authors have implemented two competing solutions which carry out their learning from user data without sending the original user data to the server and extensively discussed the unification of the two solutions. They have also designed a unified privacy loss metric for both of the solutions employing a general sample inference attack. The authors have extensively conducted various experiments for comparing both the solutions through testing them on various machine learning-based problems in certain mobile scenarios.

The authors Hu et al. [6] in their article have proposed a privacy-preserving approach to learn personalized models on distributed user data effectively while preserving the differential privacy of user data. Their approach has considered various practical matters in a distributed learning system such as user heterogeneity. They have also done a rigorous analysis of the convergence property and privacy guarantee of their approach. The experimental results of their work on realistic mobile sensing data explain and prove that their approach is quite robust to user heterogeneity and offers a good trade-off between privacy and accuracy. However, with the federated learning approaches presented in [6, 29], the reproducibility of the experiments and hence the results becomes infeasible.

Privacy preservation has found its importance in quite important domains be it location-based data hiding, preservation of credentials related to crowdsensing data, or be it the personal information of social media users. Techniques ranging from graph-based algorithms to incorporation of the newest state-of-the-art tools like deep learning, for the implementation of the same, have been in practice by researchers since the last decade. Advances have been made both in theoretical as well as application-based perspectives since the time the concept of privacy came into existence in the domain of Computer Science as a whole. HAR is a domain that still has a lot of potentials to be discovered and thus, we provide here with a framework for the same.

3 Parameters for differential privacy

Differential privacy is a framework to quantify the degree to which individual privacy is preserved while effectively releasing useful statistical information about the database which can be used for various studies. The mathematical definition of differential privacy [29] is stated as - a randomized algorithm L gives ϵ-differential privacy if for all data sets $|X-Y| \leq 1$ and any S $\subset$ Range(L), where S: All potential output of L that could be predicted.

$$Pr[L(X) \in S] \leq e^{\epsilon} Pr[L(Y) \in S] \tag{1}$$

Here, the two factors that are used to measure differential privacy are as follows.

- Accuracy is the closeness of the output of differential privacy algorithms to the original output.
- Epsilon (ϵ) in differential privacy algorithms is an assessment of privacy loss at a differential changed data.

  To simplify the definition, it can be stated that an adversary attempts to track one individual user and figure out his/her corresponding data. With some prior knowledge about the person, the probability of finding the information is $P(in)/P(out)$ of the user to be in the database than isn't, and the measure of that is ϵ in differential privacy. Thus, w.r.t our HAR problem, given the collected data as input to the classification algorithms, a privacy preserving mechanism is needed to be applied in order to keep the HAR classification model differentially private. Formulating it from (3), the differential privacy definition, $e^{\epsilon} \geq P(in)/P(out)$, where $P(in)$ in case of this instance is $F1 - Score_{perturbeddata}$ and $P(out)$ is $F1 - Score_{RealData}$. F1-score is the measure of the accuracy of the model on the given data set. Therefore, ϵ is calculated as follows.

$$\epsilon = \ln\left(\frac{F1 - Score_{perturbeddata}}{F1 - Score_{RealData}}\right) \tag{2}$$

4 Data

The accelerometer sensor readings of smartphones are collected through an Android application (such as, GSensorLogger, available in Google Play Store) and stored in CSV format. Linear acceleration along the X, Y, and Z dimensions are collected for different users carrying the smartphone in different manners. Both the hardware configuration of the smartphones and the different usage behavior are taken into account for two reasons-

- the sensor readings vary depending on these two factors.
- usage behavior should be treated as private data along with the user id.

Due to different hardware configurations, when a subject performs the same activity, sensor readings recorded by different smartphones would differ (due to sensor calibrations) [17]. Even, for the same activity performed by the same subject, if the smartphone is held at hand or kept in shirt pocket, accelerometer readings vary. For instance, if it is carried in hand, more noise would be inserted into the data. To capture this effect, data is collected from each volunteer for different smartphone configurations. The hardware configurations of the smartphones considered for data collection are summarized in Table 1. The collected accelerometer readings are labeled according to the corresponding activity performed and the (user id, usage behavior) combination is also noted as a dimension of data that should be marked as private. Different devices were used to collect the data. The accelerometer data

Table 1 Hardware specification of the devices from which the data are collected

Smartphone	Operating System	Processor	Memory	Battery
Google Pixel	Android 10	Qualcomm SDM730 Snapdragon 730G (8nm)	6GB RAM 128 GB UFS 2.1	3140 mAh
Samsung Galaxy J7	Android 8.0	Mediatek MT6757 Helio P20	4GB RAM 32GB eMMC 4.5	3300 mAh
Redmi Y1	Android 7.1.2	Qualcomm MSM8940 Snapdragon 435	4GB RAM 64 GB eMMC 5.1	3080 mAh
Nexus 4	Android 4.2	Qualcomm APQ8064 Snapdragon S4 Pro	2GB RAM 16 GB FLASH MEMORY	2100 mAh

has 3 columns X, Y, and Z representing the values of linear acceleration along the respective axes. The data is annotated by the 'activity' performed as has been observed and noted by a volunteer and also fetched by the application that reflects the state (running, walking, standing, etc.) while collecting data from a subject. Thus, during data collection from subjects, one or more volunteers were present to record the timing of each activities performed. Sometimes video recording is also done to later verify the ground truth information by the team. Person, device, and how the device is carried by the subject these information are combined to generate a "USER ID". Any of the fetched information being different resulted in a different "USER ID". So, the final prepared data that we used in this work for various operations, consisted of 1502369 rows and 5 columns.

The dataset is split into two sections; training and test dataset. The training dataset consists of 1051658 instances and the test dataset consists of 450711 instances. We have described the data collection procedure here. A detailed description of the dataset thus collected and other setup parameters is summarized in the Experimental Results section. The design of the proposed differentially private framework for activity recognition is detailed in the next section.

5 Methodology description

The proposed framework is described in the following subsections. The subsections explain the transformation, normalization, and feature extraction techniques involved in the whole framework before training the model on the data. Thereafter, the extracted features and the classifier trained on the extracted features have been explained in detail. The workflow of the proposed HAR framework has been summarized in Fig. 1. All the major steps are detailed in the following subsections.

5.1 Data preprocessing

This is the most important step in the proposed framework. The raw accelerometer instances are first filtered to clean the high frequency and low-frequency noise that gets inserted in the process of data collection. Once, the data is ready for analysis, the entire dataset is divided into two sets of (train, test) combination-one is the originally collected combination and the other one is the differentially private version. This is shown in the preprocessing block of Fig. 1. Depending on the application scenario and the kind of attacks considered, sometimes it is necessary to protect the training set while sometimes protecting the test data becomes important for user privacy. In this work, we mostly focus on the privacy of the test dataset as the privacy of the users of HAR should be given the highest priority, even over accuracy at times in order to ensure wide-scale adoption of such techniques for smart healthcare and fitness tracking. So, the collected test dataset should not be sent directly to the server, rather, the data is perturbed through our proposed technique in such a manner that the activity can be recognized precisely but the user id and behavior identification would be difficult.

5.1.1 Noise activity profiling

To guarantee differential privacy, a typical method is adding random noise to the original data for data release. There are various mechanisms of adding noise to the data to achieve

differential privacy. Noise can be introduced to the test data in the following intuitive way:

1. Linear shift, such as $[X, Y, Z] + p$, where p is the scalar quantity.
2. Parabolic shift, such as $[X, Y, Z]^p$, where p is the scalar quantity.
3. Exponential shift, such as $e^{[}X, Y, Z]$.
4. Fractional shift, such as $1/[X, Y, Z]$.

Among these listed methods, linear shift performs better than the others as it does not hamper the inter-class invariance. Samples of all activity classes are equally shifted, whereas with the other noise insertion methods there is an asymmetrical shift in the sample data. ϵ values are also calculated using (2) as in differential privacy algorithms, ϵ is a measurement of privacy loss for a differential change in data.

The noise profile can be refined further by equipping uniform noise distribution rather than taking a constant noise value for the following two reasons:

– it will limit, the horizon of the experimentation
– with a constant ϵ value, the adversary has a high chance to identify the individual's record with the released information or any auxiliary information under differential privacy.

It is known that with increase in uniform noise in the data, a decrease in ϵ value can be seen. Smaller ϵ value signifies better privacy protection.

5.1.2 Noise vector generation

A better metric in accuracy and privacy result can be achieved if the uniform scalar noise produced in the test data is treated as a vector noise. Since in the data, accelerometer readings are vectors, not scalars, so, through linear shifts, vector noise could be inserted into the data. The observation led to determining noise as a vector quantity. To serve the idea of the vector noise, uniform noise is added to the magnitude of the data, and the data is rotated randomly in any direction. To keep the magnitude the same in the uniform rotated vector noise data, the data is clipped or extended as shown in Fig. 2.

It is known that to observe a pattern in the data, the magnitude has relatively less significance in defining the property it just defines the quantity. On the other hand, the combination of direction and magnitude has more significance in determining the pattern of the data. Furthermore, it can be stated that if there are two different mechanisms M1 and M2 and both of them map for domain D to range R, for all possible subsets s of R, it holds true for

$$Pr(M1(d) \in s) \approx Pr(M2(d) \in s) \forall d \in D \tag{3}$$

Then it can be stated that the shift from one mechanism to another, will carry less significance to the prediction of the domain set. So it can be said that the behavior of the two different mechanisms is not very much different towards privacy measurement.

With that taken into account, the parameters of differential privacy data (ϵ) would further improve the results.

Summarizing the two different noise insertion mechanisms, Algorithm 1 is proposed.

Algorithm 1 $Preserve\_Privacy()$.

input : :
1 Original $TestSet, TrainSet$
output: :
2 $UpdatedTestSet$
3 **for** *each row of* $TestSet$ **do**
4 Set Amplitude = $\sqrt{X^2 + Y^2 + Z^2}$;
5 take a real no, $q \in \{+limit, -limit\}$;
6 **repeat**
7 $(X', Y', Z') \leftarrow (X + q, Y + q, Z + q)$;
8 $NewAmplitude \leftarrow \sqrt{X'^2 + Y'^2 + Z'^2}$;
9 $m \leftarrow \left(1 + \frac{NewAmplitude}{|Amplitude - NewAmplitude|}\right)$;
10 $(X'', Y'', Z'') \leftarrow (X' \times m, Y' \times m, Z' \times m)$;
11 $accuracy \leftarrow Classifier(TrainSet, TestSet)$;
12 $accuracy' \leftarrow Classifier(TrainSet, TestSet')$;
13 **if** $accuracy \approx accuracy'$ **then**
14 Calculate ϵ;
15 **else**
16 Vary $q \in \{+limit, -limit\}$;
17 **until** *convergence*;

The sensor readings along each axis has been translated by a scalar quantity in step 7 of Algorithm 1. Since the accelerometer readings are vectors, not scalars; so in the next steps (step 8-10), the data has been clipped or extended as indicated in Fig. 2. The process has been repeated until convergence as indicated by ϵ w.r.t a validation set. Features are extracted from the data. The effect of perturbation of the feature set extracted from the training data is also explored in the work. This is detailed in the following subsection.

5.2 Feature optimization

The data is segmented into overlapping windows for the extraction of features. Features better represent the patterns for different activity classes. Time-domain and frequency-domain features could be extracted that capture the temporal dependency in the accelerometer data corresponding to the various activities. The following nine features were extracted:

- MEAN(x) (Mean of X)
- MEAN(y) (Mean of Y)
- MEAN(z) (Mean of Z)
- VAR(x) (Variance of X)
- VAR(y) (Variance of Y)
- VAR(z) (Variance of Z)
- STD(x) (Standard Deviation of X)
- STD(y) (Standard Deviation of Y)
- STD(z) (Standard Deviation of Z)

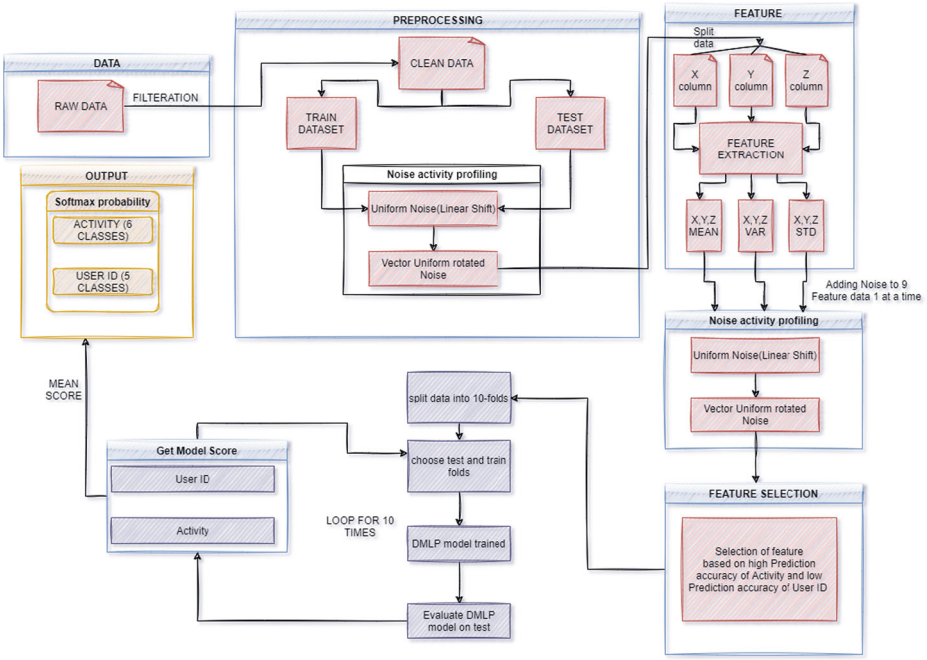


Fig. 1 Workflow of the proposed HAR framework

Feature level perturbation is also explored in the work in such a way that the important features for activity recognition are kept intact. The features are perturbed individually applying the basic principles of Algorithm 1. The effectiveness of this technique is statistically validated through adopting the cross validation technique.

5.3 Deep multi layer perceptron (DMLP)

A Deep Multi Layer Perceptron (DMLP) is a light framework with 5 hidden layers used for classification [22]. It has been used as the backbone of the whole framework. It has less complexity for pattern recognition from numerical data as compared to frameworks like Recurrent Neural Network (RNN) and Long Short-Term Memory network (LSTM) where, sequence matters [26]. Our data is not a sequential data and consists of independent and less columns, for which the DMLP is the best suite of framework, and we verified it by experimenting.

5.3.1 Definition

Feed-forward artificial neural networks(ANNs) consist of various types of networks. One such special type of network is the MLP which refers to any feed-forward ANN, that is composed of multiple layers of perceptrons with threshold activation [11].

Any basic unit of an MLP is made up of at least three layers of nodes, namely an input layer, a hidden layer, and an output layer. Other than the input nodes, every other node is a neuron that makes use of an activation function that has a nonlinear nature. Generally, it makes use of a supervised learning technique called backpropagation for the training itself.

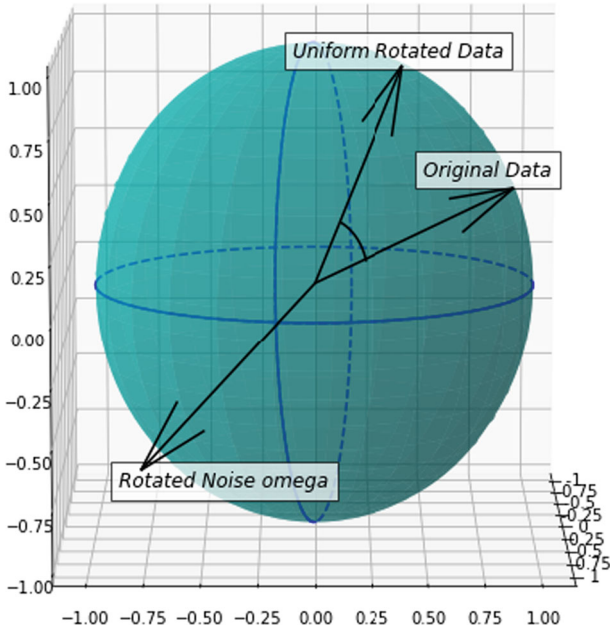


Fig. 2 Conversion of uniform noise to uniform rotated noise

Its multiple layers and non-linear activation provide the support for distinguishing data that is not linearly separable.

5.3.2 Learning

The process of learning occurs in the MLP by changing connection weights after each piece of data is processed. This processing is based on the quantity of error in the output compared to the expected output. This characterizes supervised learning and is carried out through propagating backward [4]. We can represent the degree of error in an output node j in the m the data point (training example) by $e_j(m) = d_j(m) - y_j(m)$, where d is the true(target) value and y is the value predicted by the perceptron. The weights of the node can then be adjusted based on corrections that minimize the error in the entire output. By making use of gradient descent, the change acquired in each weight is

$$\Delta w_{ji}(m) = -\eta \frac{\partial \mathbb{E}(m)}{\partial v_j(m)} y_i(m) \tag{4}$$

where y_i denotes the output of the $(j-1)^{th}$ neuron and η is the learning rate, which is accordingly selected to ensure quick convergence of the weights to a change, without oscillations.

The derivative to be calculated depends on the induced local field v_j, which itself varies. Proofs exist about the fact that for an output node this derivative can be simplified to

$$-\frac{\partial \mathbb{E}(m)}{\partial v_j(m)} = e_j(m)\beta'(v_j(m)) \tag{5}$$

where β' is the derivative of the activation function, which itself is non-varying in nature [14]. This analysis increases in complexity for the subsequent updating of weights in a hidden node. This depends on the updating of weights in the kth nodes, which denote the output layer. So to change the hidden layer weights, the weights of the output layer change in accordance with the derivative of the activation function. So, this part of the algorithm denotes a backpropagation of the activation function.

6 Experimental results

The experiments are carried out using the Scikit Learn toolkit [13] of the Python programming language. Google colab platform has been utilized for the experimentation. A graphics package in Python [7] is utilized for the data plots. The dataset details and the experimental setup are summarized in Table 2. As shown, data is collected from 8 users of the age group 20-32 years. Temporal features, such as, mean, standard deviation, variance, and so on are extracted as the features along the (x,y,z) dimensions. It is found that when Algorithm 1 is applied to the data and noise is inserted to the instances of standard deviation along the z-axis, a perfect balance could be achieved between the privacy and activity recognition accuracy. The following experiments are conducted to validate our observation and evaluate the performance of the algorithm.

6.1 Before feature extraction

The first experiment is conducted to find the relationship between privacy loss and the activity recognition accuracy for the proposed data perturbation techniques. ϵ is a measure of the privacy of the dataset. The lesser the values of ϵ, the better is the privacy. It is computed for the various levels of noise insertion and the corresponding accuracy is also observed in Fig. 3. It can be observed that uniform rotated noise works better than uniform noise as it better preserves the class invariance relationship for activity recognition.

It is also important to observe if, for a given perturbed dataset, the privacy-preserving attribute, that is the user details can be preserved while securing good recognition performance for HAR. To explore this, an experiment is conducted by varying noise levels and the results are plotted in Fig. 4. As shown in the figure, the proposed method is found not to hamper the HAR data patterns. However, a drop in accuracy for user id prediction has been

Table 2 Summary of the experimental setup

Attributes	Values
Devices	4
Data sampling rate	50 samples/sec
Filtering method	Butterworth Filter, Median Filter
Window type and size	2 sec sliding window with 1 sec overlapping window
Position	Shirt Pocket, Right Pocket of Trouser, Hand
Number of users	8
Number of activity classes	6
Activity classes	Sit, Stand, NWalk, SWalk, BWalk, Jogging
Number of features	9

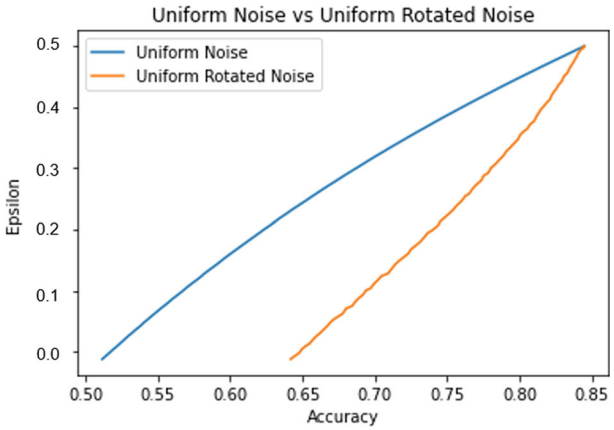


Fig. 3 Epsilon vs prediction accuracy for Uniform noise inserted to the accelerometer data

observed indicating that the privacy of user id is preserved well. This is investigated in more detail through the following experimentation.

The effect of the perturbation introduced through the proposed algorithm is analyzed in Fig. 5. The data from 5 users have been shown who performed 6 different activities. As can be observed from Fig. 5a and c, though the data has been perturbed, the class boundaries are retained. Hence, if we compare the activity prediction performance shown in Fig. 5c and e, all the different classes are well predicted by the classifier. The prediction performance is as good as the prediction of data without noise. The original data without noise is shown in Fig. 5a and its prediction performance is shown in Fig. 5g. However, if the user ID data is observed, as indicated by Fig. 5b and h, the user id can be well predicted through the classifier, thus posing privacy concerns. However, the result of perturbation on user ID is shown in Fig. 5d and its prediction performance is shown in Fig. 5f. If the scattered points are followed, change in the shades in the two figures for the same points indicate that the points are

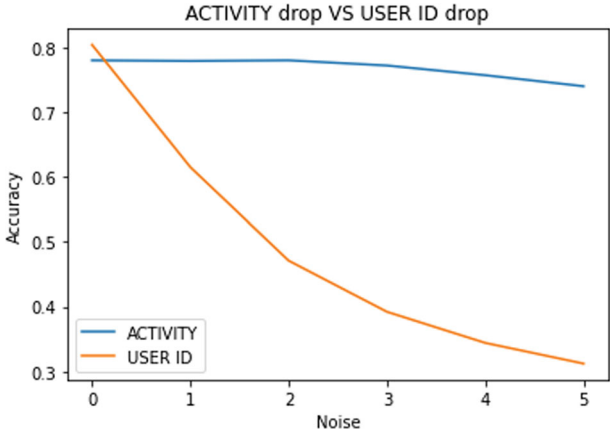
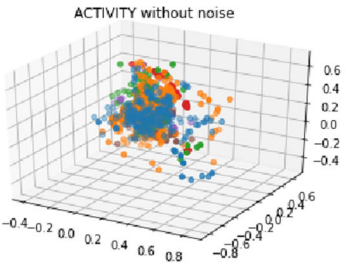
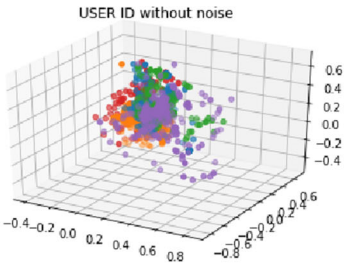
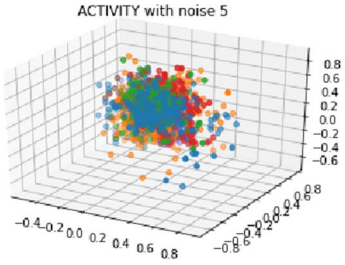


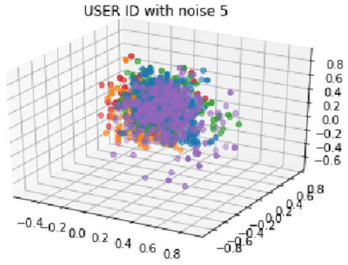
Fig. 4 Prediction accuracy for both ACTIVITY and USER ID prediction when noise is inserted to the accelerometer data



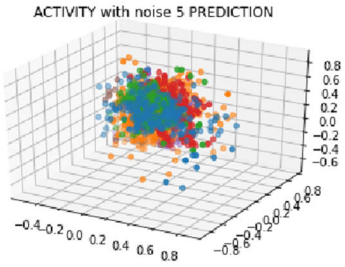


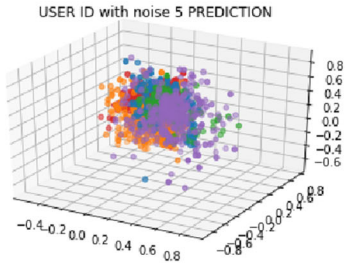
(a) Activity wise data points without noise (b) User ID wise data points without noise



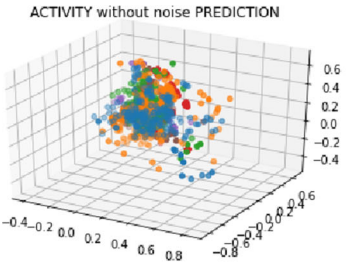


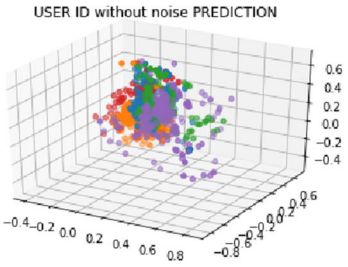
(c) Activity wise data points with noise (d) User ID wise data points with noise





(e) Predicted Activity wise data points with noise (f) Predicted User ID wise data points with noise





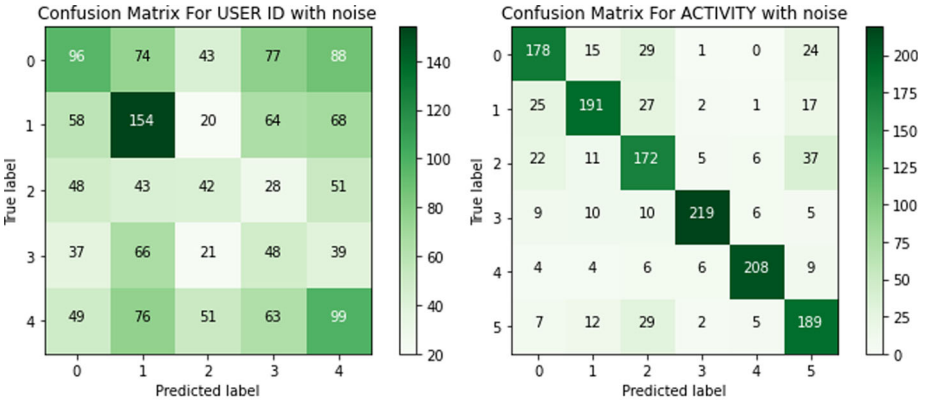
(g) Predicted Activity wise data points without noise (h) Predicted User ID wise data points without noise

Fig. 5 Data points indicating the prediction performance

wrongly predicted. So, for the input data pattern, both user ID and activity classes can be distinguished well through supervised learning as is evident from Fig. 5a, b, g, and h. However, after inserting noise following the proposed Algorithm 1, activity prediction performs better than user ID identification. This proves the effectiveness of our algorithm in preserving the differential privacy without much affecting the activity recognition performance.

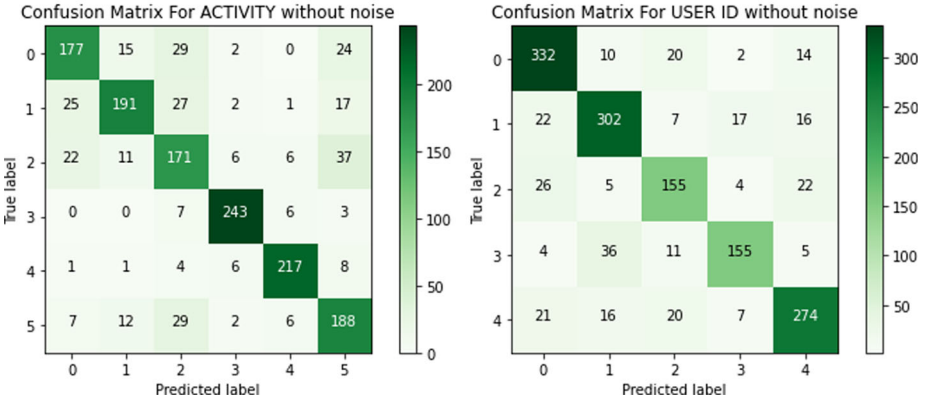
The confusion matrices as shown in Fig. 6 also support the statement.

So, it can be seen from Fig. 6 that for activity, both the matrices (with and without noise) have support concentrated almost along the diagonal. The same goes for user id results without noise. But the matrix with noise for user id has highly unevenly distributed support. From this observation, it can be inferred that the noise insertion affects the model performance quite drastically (see Fig. 6a) and makes it quite impossible to backtrack and identify the user id from their activity behavior, without affecting HAR task (see Fig. 6b) by much considerable amount, which is the main goal of our proposed work.



(a) Confusion Matrix for predicted USER ID with data perturbation

(b) Confusion Matrix for predicted ACTIVITY with data perturbation



(c) Confusion Matrix for predicted ACTIVITY without data perturbation

(d) Confusion Matrix for predicted USER ID without data perturbation

Fig. 6 Confusion matrix indicating the prediction performance

The proposed algorithm is not tied to one particular classifier but privacy is preserved and still data can be analyzed subject to other classification mechanisms as well. The results of the comparison with other classifiers have been summarized in Table 3. Commonly used deep learning classifiers, such as, 1D Convolutional Neural Network (CNN) and LSTM as well as machine learning classifier, Decision Tree have been applied. The results indicate the effectiveness of the proposed scheme in attaining good accuracy for activity recognition. Even LSTM is working fairly well for not so large dataset as collected from the volunteers.

Experiments are conducted for exploring the applicability of the proposed framework on benchmarks datasets. Two benchmark datasets are considered-(i) UCI HAR[2]; and (ii) WISDM[3]. In both the datasets raw data are provided for accelerometer values along the X, Y, and Z axes for basic activities (Walking, Jogging, Sitting, Standing, Upstairs, Downstairs). Even a few time domain and frequency domain features are also provided. Data is collected from 30 subjects for the UCI HAR dataset while data from 51 subjects have been reported in WISDM dataset. The dataset reports accelerometer and gyroscope readings for 18 activity classes including sit, stand, walk, jog, and so on. In WISDM dataset, user details for the data instances are provided. When the proposed algorithm is applied on both the datasets, the results are shown in Table 4. It can be observed that privacy could be preserved while retaining appreciable accuracy for activity classification. This indicates the generality of the approach for different datasets of the HAR domain.

6.2 Cross validation after feature extraction

k-fold cross-validation is a statistical performance measure that provides relevant performance results to cover many variations between the train and test conditions. Here, the dataset is divided into folds and ensuring that each fold is used as a testing set at some point. To estimate how accurately the predictive model will perform in practice, we have used k-fold (k=10 here) cross-validation, the results of which are shown in Table 5 and Fig. 8. So, our data set is divided into 10 groups. At each iteration, 1 group is set as test data set while the remaining k-1 groups as a train data set.

For cross validation, these steps are followed:

1. One feature at a time is perturbed.
2. All rows in the dataset are shuffled.
3. The dataset is divided into 10 folds.
4. One fold is used at a time as the test set and the combination of 9 others as a training set.

The results obtained through the above steps is summarized in Table 5. The result shows better performance when standard deviation(z) feature is perturbed. Standard deviation is a statistical term used to measure the amount of variability or dispersion around an average, which technically suffices to a measure of volatility. The Z-axis is going through the screen of the smartphone perpendicularly.

The sagittal axis of the human body is very much similar to the Z-axis of a smartphone. Now, Fig. 7 shows two use cases - one is when this smartphone is kept in a shirt pocket and another when it is kept in a pant pocket.So, it can be seen that when the smartphone is kept in a shirt pocket, the angle between the Z-axis of the smartphone and the sagittal axis of the human body is very less (about 0 degrees). But when the smartphone is kept in a pant pocket,

[2]https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones
[3]https://www.cis.fordham.edu/wisdm/dataset.php

Table 3 Comparison result on different classification models

Model description	Model's Accuracy on USER ID	Model's Accuracy on ACTIVITY
DMLP+Differential Privacy	49.781%	96.098%
CNN+Differential Privacy	49.982%	90.451%
LSTM+Differential Privacy	49.014%	89.667%
DT+Differential Privacy	50.411%	89.714%

Decision Tree

CNN : 1D (verbose, epochs, batch_size = 0, 10, 32), Activation fun = "Relu", Backpropagation = NO

LSTM : 1 hidden layer, (verbose, epochs, batch_size = 0, 15, 64), Activation fun = "Relu", Backpropagation = NO

Table 4 Classification result on different datasets

Dataset	Model's Accuracy on USER ID	Model's Accuracy on ACTIVITY
UCI HAR	–	90.787%
WISDM HAR	50.031%	92.106%
Collected Data	49.781%	96.098%
Collected Anonymous Data		94.314%

Table 5 Prediction performance when noise is inserted to one of the features

Noise inserted feature	Accuracy of predicted activity	Accuracy of predicted user Id
None	0.96	0.95
MEAN(x)	0.22	0.23
MEAN(y)	0.31	0.23
MEAN(z)	0.30	0.23
VAR(x)	0.29	0.24
VAR(y)	0.23	0.23
VAR(z)	0.31	0.24
STD(x)	0.28	0.23
STD(y)	0.38	0.23
STD(z)	0.41	0.23

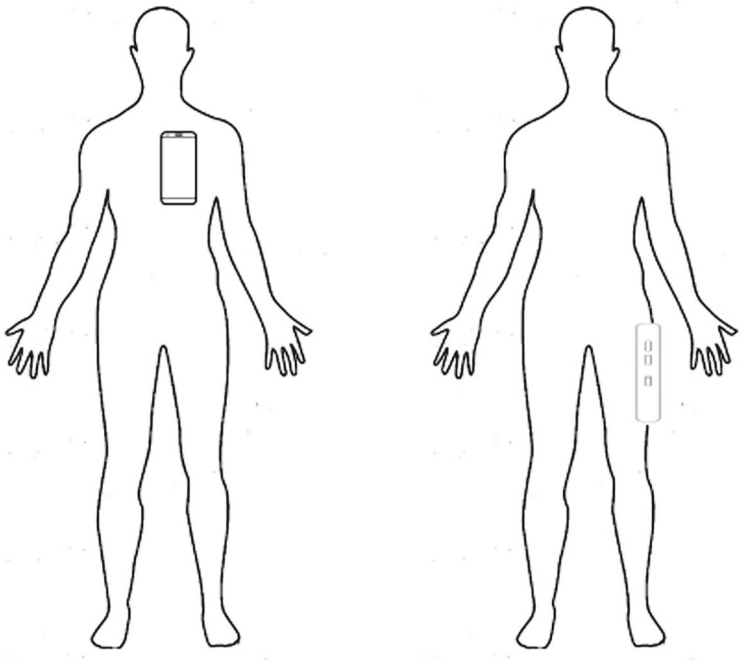


Fig. 7 Two different smartphone positions in daily life

that angle is about 90 degrees. So, STDZ or the standard deviation of the Z-axis values of smartphone accelerometer data has a great influence in determining usage behavior. Thus, when this feature is perturbed, the model is confused to determine where the smartphone is kept. However, it does not have much impact in determining the individual activity classes. This is also evident from Fig. 8. Here, the prediction performance is gradually found to stabilize as X and Y axis features are found to be more significant for activity prediction.

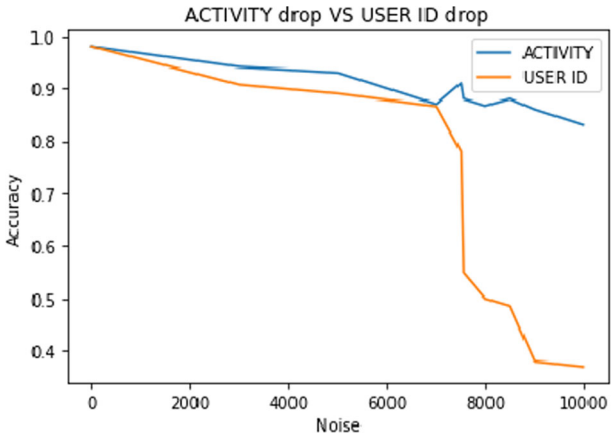


Fig. 8 Accuracy vs Noise for both ACTIVITY and USER ID prediction based on featured data when STDZ is perturbed for 100 epochs

Table 6 Result of Activity and User ID prediction for unknown test data set

SUBJECT 1	Activity	Accuracy
	SIT	0.9198
	STAND	0.9867
	WALK	0.8101
	Total	0.9188
SUBJECT 3	**Activity**	**Accuracy**
	SIT	1.0000
	STAND	0.9896
	WALK	0.8935
	Total	0.9609
SUBJECT 1 + SUBJECT 3 Total		0.943
SUBJECT 3 Usage Behavior Prediction		
Actual Smartphone Position		**Accuracy**
RPT		0.4978

Holding the phone at hand also provides similar observation as found for the other two use cases of shirt pockets and pant pockets. However, with hand movement, the data becomes more noisy.

Finally, an experiment is conducted with a slightly different setup. Here, the feature set extracted from the training data is perturbed to check the effectiveness of the proposed scheme against the inversion attacks. The other motivation is to check whether the perturbed dataset still remains workable for activity analysis of unknown test data. So, data collected from 2 new users are used to form the test set. The prediction results are reported in Table 6. As can be observed from the table, static activities are hardly affected through the perturbation. DMLP classifier provides appreciable accuracy for dynamic activities as well. Since the users are new, so, user identification cannot be done but usage beahvior prediction is investigated. It could be found that even for the unknown users also, the perturbation scheme is able to provide strong privacy thus poor accuracy as compared to the activity recognition performance.

7 Conclusion

In this paper, the problem of privacy-preserving HAR for smartphone users has been investigated. The problem is perceived from the differential privacy aspects as differential privacy is a strong notion of privacy that does not assume any specific behavior of the adversary. The accelerometer instances are collected from 8 subjects for various activities to form the dataset. It is found that through deep learning techniques, it is possible to identify a user, even his/her usage behavior, such as, how the smartphone is carried (such as right pant pocket, shirt pocket, hand) from the accelerometer trails. Thus, inversion attacks are possible on such datasets. In order to preserve the privacy of the users, perturbation techniques are explored here for both the instance space and feature space in a way that the data remains to be analyzed for HAR but no private information about the user could be predicted from it. The differential privacy measures and their effect on the corresponding activity measure are also investigated in the work. The experiments revealed that with the proposed approach,

HAR accuracy remains to be about 96% while the user prediction accuracy drops to around 49%. The methodology discussed in the paper can be used in various data intensive applications where the institution/companies can analyze activitities such as, running speed or step count (as in StepSetGo[4])) but will not be able to derive other user specific confidential information.

A limitation of the approach could be that the proposed technique is based on the domain knowledge of accelerometer data and smartphones. So, in future we plan to modify the approach in order to incorporate other sensing modes of HAR and thus, make it applicable for sensor fusion based HAR applications as well. The effectiveness of the proposed approach against the popular inference attacks is also planned to be explored.

Declarations

Conflict of Interests The authors declare that there is no conflict of interest.

References

1. Agarwal R, Hussain M (2021) Generic framework for privacy preservation in cyber-physical systems. In: Progress in advanced computing and intelligent engineering, Springer, pp 257–266
2. Boutsis I, Kalogeraki V (2013) Privacy preservation for participatory sensing data. In: 2013 IEEE International conference on pervasive computing and communications (PerCom), IEEE, pp 103–113
3. Clauset A, Moore C, Newman ME (2008) Hierarchical structure and the prediction of missing links in networks. Nature 453(7191):98–101
4. Dong S, Wang P, Abbas K (2021) A survey on deep learning and its applications. Comput Sci Rev 40:100,379. https://doi.org/10.1016/j.cosrev.2021.100379
5. Fredrikson M, Jha S, Ristenpart T (2015) Model inversion attacks that exploit confidence information and basic countermeasures. In: Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, pp 1322–1333
6. Hu R, Guo Y, Li H, Pei Q, Gong Y (2020) Personalized federated learning with differential privacy. IEEE Internet Things J. 7(10):9530–9539
7. Hunter JD (2007) Matplotlib: A 2d graphics environment. Comput Sci Eng 9(3):90–95. https://doi.org/10.1109/MCSE.2007.55
8. Ji S, Mittal P, Beyah R (2016) Graph data anonymization, de-anonymization attacks, and de-anonymizability quantification: A survey. IEEE Communications Surveys & Tutorials 19(2):1305–1326
9. Kairouz P, Oh S, Viswanath P (2014) Extremal mechanisms for local differential privacy. Advances in Neural Information Processing Systems 4(January):2879–2887
10. Kasiviswanathan SP, Nissim K, Raskhodnikova S, Smith A (2013) Analyzing graphs with node differential privacy. In: Sahai A (ed) Theory of cryptography. Springer, Heidelberg, Berlin, pp 457–476
11. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
12. Liu AX, Li R (2021) Predictable privacy-preserving mobile crowd sensing. In: Algorithms for data and computation privacy, Springer, pp 313–346
13. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. J Mach Learn Res 12:2825–2830
14. Rojas R (1996) Neural networks: A systematic introduction springer. Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-61068-4
15. Ryoo M, Rothrock B, Fleming C, Yang HJ (2017) Privacy-preserving human activity recognition from extreme low resolution. In: Proceedings of the AAAI conference on artificial intelligence, vol 31
16. Saha J, Chowdhury C, Ghosh D, Bandyopadhyay S (2020) A detailed human activity transition recognition framework for grossly labeled data from smartphone accelerometer. Multimed Tools Appl 1–22

[4] https://www.stepsetgo.com/

17. Saha J, Chowdhury C, Roy Chowdhury I, Biswas S, Aslam N (2018) An ensemble of condition based classifiers for device independent detailed human activity recognition using smartphones. Information 9(4):94
18. Sahnoune Z, Aïmeur E (2021) Deloc: A delegation-based privacy-preserving mechanism for location-based services. Int J Mob Commun 19(1):22–52
19. Samarah S, Al Zamil MG, Aleroud AF, Rawashdeh M, Alhamid MF, Alamri A (2017) An efficient activity recognition framework: Toward privacy-sensitive health data sensing. IEEE Access 5:3848–3859
20. Shun Z, Benfei D, Zhili C, Hong Z (2021) On the differential privacy of dynamic location obfuscation with personalized error bounds. arXiv:210112602
21. Song C, Ristenpart T, Shmatikov V (2017) Machine learning models that remember too much. In: Proceedings of the 2017 ACM SIGSAC Conference on computer and communications security, pp 587–601
22. Stamate C, Magoulas G, Kueppers S, Nomikou E, Daskalopoulos I, Luchini M, Moussouri T, Roussos G (2017) Deep learning parkinson's from smartphone data. In: 2017 IEEE international conference on pervasive computing and communications (PerCom), pp 31–40. https://doi.org/10.1109/PERCOM.2017.7917848
23. Sweeney L (2002) Achieving k-anonymity privacy protection using generalization and suppression. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(05):571–588
24. Tran AT, Luong TD, Karnjana J, Huynh VN (2021) An efficient approach for privacy preserving decentralized deep learning models based on secure multi-party computation. Neurocomputing 422:245–262. https://doi.org/10.1016/j.neucom.2020.10.014. https://www.sciencedirect.com/science/article/pii/S0925231220315095
25. Vecchio A, Mulas F, Cola G (2017) Posture recognition using the interdistances between wearable devices. IEEE Sensors Letters 1(4):1–4
26. Wan S, Liang Y, Zhang Y, Guizani M (2018) Deep multi-layer perceptron classifier for behavior analysis to estimate parkinson's disease severity using smartphones. IEEE Access 6:36,825–36,833. https://doi.org/10.1109/ACCESS.2018.2851382
27. Wang W, Zhang Q (2016) Privacy preservation for context sensing on smartphone. IEEE/ACM Trans Networking 24(6):3235–3247
28. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ (2019) Paga: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 20(1):1–9
29. Zheng H, Hu H, Han Z (2020) Preserving user privacy for machine learning: Local differential privacy or federated machine learning? IEEE Intell Syst 35(4):5–14