

# A Systematic Literature Review on Wearable Health Data Publishing under Differential Privacy

Munshi Saifuzzaman<sup>a,1</sup> , Tajkia Nuri Ananna<sup>a,1</sup> , Mohammad Jaber Morshed Chowdhury<sup>b,3</sup> , Md Sadek Ferdous<sup>c,2</sup> , Farida Chowdhury<sup>d,1</sup> 

<sup>1</sup> Shahjalal University of Science and Technology, Kumargaon, Sylhet-3114, Bangladesh

<sup>2</sup> BRAC University, Dhaka-1212, Bangladesh

<sup>3</sup> La Trobe University, Bundoora, VIC, 3086, Australia

Received: date / Accepted: date

**Abstract** Wearable devices generate different types of physiological data about the individuals. These data can provide valuable insights for medical researchers and clinicians that cannot be availed through traditional measures. Researchers have historically relied on survey responses or observed behavior. Interestingly, physiological data can provide a richer amount of user cognition than that obtained from any other sources, including the user himself. Therefore, the inexpensive consumer-grade wearable devices have become a point of interest for the health researchers. In addition, they are also used in continuous remote health monitoring and sometimes by the insurance companies. However, the biggest concern for such kind of use cases is the privacy of the individuals. There are a few privacy mechanisms, such as abstraction and  $k$ -anonymity, are widely used in information systems. Recently, Differential Privacy (DP) has emerged as a proficient technique to publish privacy sensitive data, including data from wearable devices. In this paper, we have conducted a Systematic Literature Review (SLR) to identify, select and critically appraise researches in DP as well as to understand different techniques and exiting use of DP in wearable data publishing. Based on our study we have identified the limitations of proposed solutions and provided future directions.

**Keywords** Wearable, Health Data, Real-time Health Data, Privacy, Differential Privacy.

## 1 Introduction

Recent advances in wearable and smart technology, and the rapid adoption of wearable devices and smartphones makes them an important source of information for healthcare and medical research. The availability of these devices and the types of parameters they can measure is rapidly increasing. Real-time participant-generated physiological data can enable large scale observational studies of health conditions, provide better insights into the medical conditions of individuals, and help streamline clinical trial processes in medical research.

The researchers at IBM Watson stated than an average person possibly generates more than one million gigabytes of health-related data across his or her lifetime [1]. These data are mostly physiological and personally identifiable data such as heart rate, blood pressure, respiratory rate, disease symptoms etc. These data generated from smart wearable healthcare devices have become a blessing for the modern healthcare. Whether remotely monitoring patients or keeping track of physical condition and fitness, these data have

the potential to transform the healthcare sector. The main difference between traditional healthcare data and the wearable device generated data is that data from wearables are of continuous nature which are updated dynamically and often can be temporally correlated. These data are used by medical professionals for continuous monitoring, researchers, analysts, insurance companies or sometimes even in health surveys [2].

However, data privacy is a major concern for health data [3]. The need to ensure privacy and trust when sharing an individual's health data is particularly critical given the sensitive nature of health data and its protection by legislation (e.g. [4] [5]). Breaches of such privacy are not uncommon [6]. The level of trust between participants in an open health data marketplace will be lower than, say, a hospital's clinical practice where the doctors and patients are known to each other. Therefore, appropriate mechanisms need to be established to ensure data security and integrity, and to build trust among the participants.

For preserving the privacy of sensitive data, many solutions have been proposed such as cryptography [7, 8], blockchain

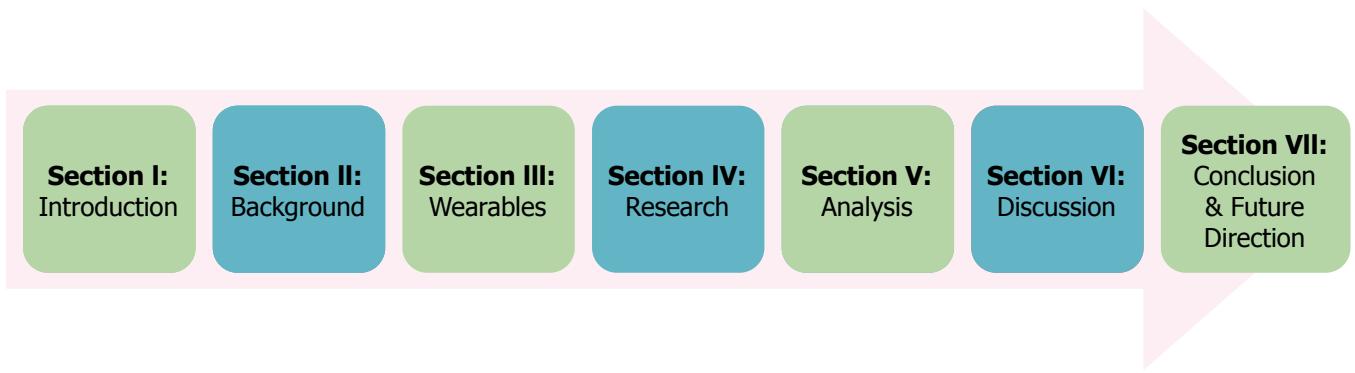


Fig. 1: Organization of the paper

[9–13], anonymization [14–18], privacy policy [19], and access control [20, 21]. All of these techniques and mechanisms have limitations, specially for publishing real-time dynamic data. A recent privacy technique, called Differential Privacy by Dwork [22] has revolutionized researches in the privacy domain. DP ensures the privacy of the individuals in a way that the presence or absence of any individual in the published dataset cannot be discovered. This reduces the risk of privacy leakage of sensitive real-time data to a great extent [23]. Therefore, it is used by different technology giants like Facebook, Google and Uber to protect the privacy of their user [24–26].

One of the biggest challenges in real-time data is the high dimensional temporal correlation between data. Many researchers have found differential privacy suitable for preserving privacy in real-time health data and claimed that these solutions have advantages over existing methods. This proves that differential privacy is a fruitful mechanism and provides a more practical way for preserving privacy of real-time health data. To the best of our knowledge, there is no dedicated survey, traditional literature review, or a Systematic Literature Review (SLR) for privacy-preserving wearable physiological data publishing using DP. This motivates our work in this paper.

We have performed an SLR on wearable data publishing (which generates data in a continuous manner) under differential privacy in the period from 2007 to April 31, 2020. We have come up with a holistic view of preserving privacy of wearable physiological data according to the existing literature. By performing a systematic mapping, we have analyzed the techniques, their use cases, datasets, experiment scenario and limitations of the existing solutions. We have categorized the research papers mainly into three major parts: *Physiological, Real-time, and Others*. We have explored and analyzed how the research community have addressed them, how they have contributed by approaching different types of techniques, what experimental procedure they have considered and what limitations they have summed up.

We have illustrated the structure of this paper in Fig. 1. In Section 2, we have narrated necessary mathematical concepts of differential privacy, and their basic mechanisms. In Section 3, we have discussed about wearable devices, types of data they generate and difference between wearable health data and traditional health data. We have explained the systematic literature review process in Section 4 and provided analysis in Section 5. Section 6 has provided a brief discussion. Finally, we have concluded with future directions in Section 8.

## 2 Background

### 5

In this section, we have provided the definition of differential privacy and its different variants. We have also discussed about different relevant concepts related to differential privacy.

#### 2.1 Differential Privacy

Differential privacy is the process of providing privacy of database in such a way that it should not reveal any Personal Identifiable Information (PII) about any individual for any query. In other words, nobody can ascertain the participation or non-participation of any individual in any dataset. The final result will not be affected by the presence or absence of any individual. Fig. 2 shows a simple visualization of differential privacy. In the upper database, Munshi, Tajkia, and Bob have contributed their respiratory data. Eve (adversary) wants to find out Bob's respiratory rate. If we delete Bob's respiratory data, the probability of finding Bob's data in the database before and after deletion will be identical. Which means that, Eve can not ascertain whether or not Bob is included in the dataset, let alone the contents of his data. Hence, Bob's privacy is preserved.

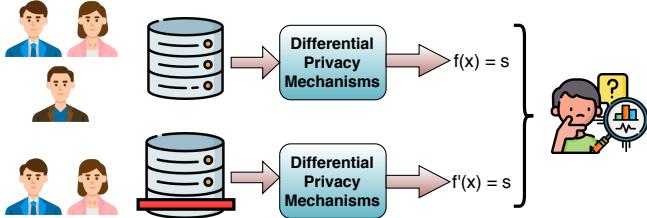


Fig. 2: Differential privacy

### 2.1.1 Definition

A randomized mechanism  $M$  gives  $(\epsilon, \delta)$ -DP for every set of outputs  $S$ , and for any neighbouring datasets (datasets that differ in only one value) of  $D, D'$  if  $M$  satisfies Eq. 1 [27]:

$$\frac{\Pr[M(D) \in S]}{\Pr[M(D') \in S]} \leq e^\epsilon + \frac{\delta}{\Pr[M(D') \in S]} \quad (1)$$

This is known as *approximate differential privacy*. If  $\delta = 0$ , then Eq. (1) shows the ratio between the probability of the output being into dataset  $D$  and  $D'$  becomes less than or equal to  $e^\epsilon$ . This is known as *pure differential privacy*. If two datasets differ with  $c$  values then the ratio becomes less than or equal to  $e^{\epsilon c}$ . This is known as *group privacy*.

The mechanism  $\epsilon$  and  $M$  are the main actors here.  $\epsilon$  is the balance between privacy loss and maximizing utility.

1.  $\epsilon = 0$  leads to complete privacy but lack of utility.
2.  $\epsilon <= 1$  leads to less privacy but higher utility.

$M$  decides how much noise (i.e. a calibrated value used to anonymize data) will be added and what type of query is being served.

### 2.1.2 Illustration of Differential Privacy

Let us consider a child named John. There is a large amount of data of John which has been generated during John's lifetime. This data is managed by John's parents. One example of such data is John's preference: which veggies John dislikes? However, he is a bit ashamed about his preference and hence, may not wish for everyone to know this. The ability to keep this type of secret is referred to as privacy. However, the man who prepares John's lunch at his daycare may wish to know that some of the children in the group dislike carrots! He is not required to know whether or not John is one of these children. It is sufficient if he is aware that there are perhaps four or five children who dislike carrots. This is referred to as differential privacy. Now, if the cook asks John whether he likes carrots or not, instead of giving a direct answer, he could utilize an approach which simulates differential privacy. In this approach, John will provide an indirect answer. More specifically, he will answer it through an outcome of some random process, e.g. a coin flip. The

cook cannot see the result of this random process, this is essential. John flips the coin. The process is shown Fig. 3.

1. If it is head, John will provide the true answer.
2. If it is tail, John will flip the coin again.
  - If it is head, John will say yes, regardless of what the true answer is.
  - If it is tail, John will say no, regardless of what the true answer is.

This is an example of plausible deniability which refers to an individual's ability to deny anything since there is no concrete proof to show him right or wrong. Continuing with the example, the randomness of flipping a coin allows John to be protected with plausible deniability as it is plausible for him to deny the answer based on the outcome of coin flipping. This randomized response process is actually differentially private process. Although the algorithms for differential privacy are much more complex, the principle remains the same. By making it unclear whether or not each response is legitimate, or even by altering replies arbitrarily, these algorithms can assure that regardless of how many queries are sent to the database, no one can be identified concretely.

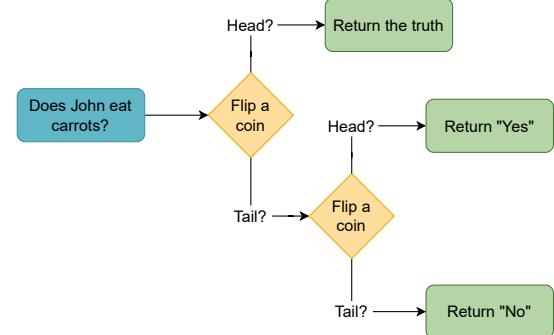


Fig. 3: Flow diagram of the Differential privacy algorithm

### 2.1.3 Local Differential Privacy (LDP)

LDP [28, 29] is a slightly different method to achieve differential privacy. In other words, it is used to provide differential privacy locally. Differential privacy was designed for the purpose of sharing data, whereas LDP protects the process of data collection by maintaining individual privacy. In this case, a user instead of giving true data directly to the aggregator (an entity that collects as well as aggregates data from different sources and anonymize them), they add noise to their individual data first, then send the noisy data to the aggregator. This ensures individual privacy. Fig. 4 shows the differences between DP and LDP.

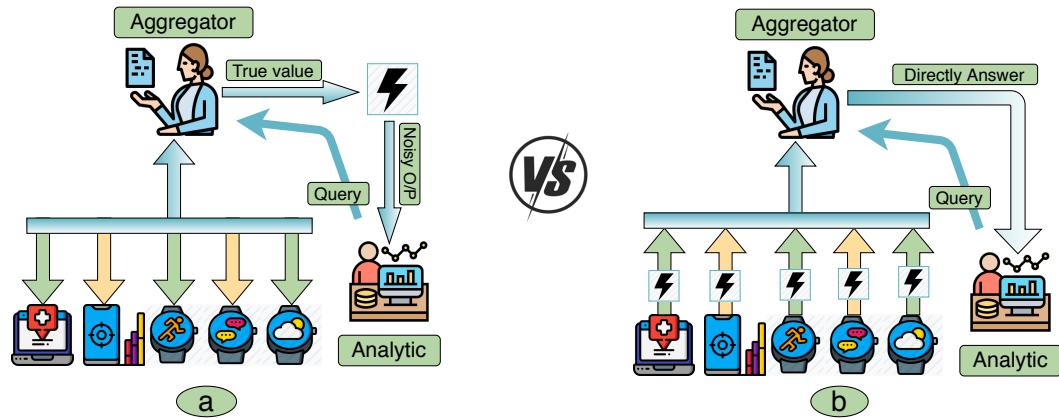


Fig. 4: Differences between (a) DP and (b) LDP

#### 2.1.4 Sensitivity

Sensitivity is the maximum difference on query results between neighbouring datasets and defined as  $\Delta f$  [30]. Suppose there is a query "How many people in the database have the property  $P$ ?". In this scenario, the presence or absence of an individual will change the result to a maximum value of only 1. So the sensitivity of this dataset is just 1. There are two types of sensitivity in differential privacy, namely *Local* sensitivity and *Global* sensitivity.

#### 2.1.5 Privacy Budget

$\epsilon$  is known as the privacy budget which controls the privacy guarantee level of any mechanism  $M$  [27]. The main responsibility of the privacy budget is to maintain the balance between privacy loss and utility maximization. Smaller  $\epsilon$  (i.e more noise) ensures stronger privacy. But due to smaller epsilon the data can lose its utility and vice-versa. So, it is important to find and maintain the balance between privacy loss and utility maximization. Fig. 5 shows a visual representation of this problem. As shown in figure, the more noise is added to the face image, the more anonymous it gets. But at the same time, with more anonymization the image becomes less useful. Similarly, less noise preserves utility of the image but does not provide any considerable privacy.

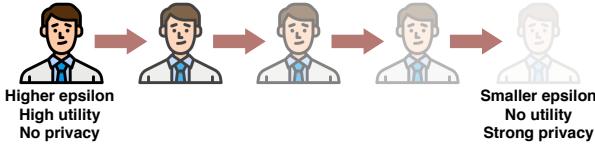


Fig. 5: Effect of privacy budget

## 2.2 Mechanisms of Differential Privacy

Differential privacy can be achieved in two different ways, namely interactive and non-interactive. In interactive way, the system response to each query individually until total privacy budget is consumed. Whereas in non-interactive way, the data curator (who maintains and manages metadata) either evaluates and brings out statistics or discloses raw data anonymously. All the query responses are given at a time. In addition, DP uses different types of mathematical and statistical model for data perturbation based on the types of data such as numeric data and non-numeric data [27].

1. For **numeric queries**, Laplace and Gaussian mechanisms are more suitable.
2. For **non-numeric queries**, Exponential mechanism is more suitable.

### 2.2.1 Laplace Mechanism

The *laplace mechanism* is the procedure of adding *laplace noise* to the query result [30]. The noise is sampled from the *laplace distribution* [27]. Eq. 2 shows the probability density function for *laplace distribution* which is centered at 0 with scale  $b$ :

$$\text{Lap}(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right) \quad (2)$$

The *laplace mechanism* uses  $l_1$ -sensitivity (magnitude by which a single individual's data can change) and the variance of this distribution is  $\sigma^2 = 2b^2$ .

### 2.2.2 Gaussian Mechanism

In *gaussian mechanism*, gaussian noise is added to the function [27]. Rather than scaling to  $l_1$ -sensitivity, curator scales

it to the  $\ell_2$ -sensitivity. Eq. 3 shows the mechanism of adding *gaussian noise* to the results.

$$M(D) = f(D) + N(0, \sigma^2) \quad (3)$$

Where  $\sigma = \Delta_2 f \sqrt{2 \ln(2/\delta)} / \varepsilon$ . And  $N(0, \sigma^2)$  is the added *Gaussian noise*.

### 2.2.3 Exponential Mechanism

The *exponential mechanism* [27] is used in case of non-numeric attributes because both the *laplace mechanism* and *gaussian mechanism* cannot deal with non-numeric attributes. In this case, the quality of an outcome is measured using a *score function*. The score function  $q(D, \phi)$  represents how good an output  $\phi$  is for the dataset  $D$ . Eq. 4 represents the equation of exponential mechanism.

$$M(D) = \{ \text{return } \phi \text{ with the probability} \} \propto \exp\left(\frac{\varepsilon q(D, \phi)}{2\Delta_q}\right) \quad (4)$$

Where  $\Delta_q$  represents the sensitivity of score function  $q$ .

## 3 Wearables

Wearable devices can collect real-time data such as spatio-temporal data, trajectory data, location data but most importantly physiological data. These can track activities and remotely monitor a patient's condition. Consumer grade wearable trackers can track fitness and biosensors can collect biological data. Example of other wearables are smart footwear which includes smart shoes, socks, insoles and gloves [31], smart jewellery that includes smart ring, smart bracelet even smart band, smart eye wear, glucose monitoring device, blood pressure monitor, body mounted sensor and biosensor.

### 3.1 Wearable Device Architecture

The architecture of wearable devices' data exchange consists of three major components: 1) Wearable device 2) Smartphone and 3) Server/cloud server. Fig. 6 illustrates the wearable scenario, demonstrating the data generation and sharing process.

The wearable device collects data using sensors attached to it and transmits that data to the user's smartphone. Bluetooth is used to transmit data between the wearable device and the smartphone. These sensors data are transmitted to the smartphone continuously and in real time. User applications on the smartphone enable the user to monitor the data collected by the wearable. After that, the data stored in the smartphone are transmitted to the remote server/cloud server via mobile network or WiFi. These server-stored data are shared with healthcare providers, health researchers, or immediate family members based on the user's preferences.

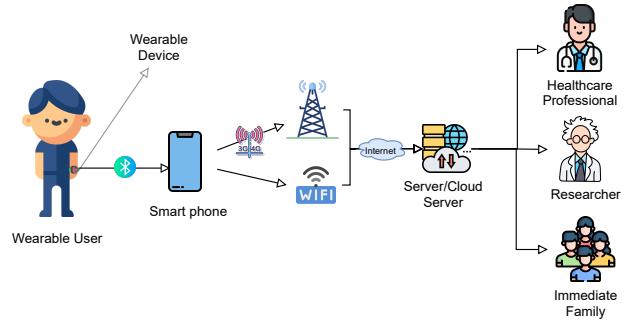


Fig. 6: Architecture of a wearable device

### 3.2 Types of Physiological Data

Wearable device technology has become a crucial tool in the world of healthcare. These devices collect real-time physiological data continuously. Individuals can track their health condition and physical activities. Doctors can also monitor their patients remotely without having the patient to visit him/her in person or they can take necessary steps in case of emergency.

Some of the wearable devices are medical grade and some are customer grade devices such as Apple watch or Fitbit [32]. In recent times, the customer grade devices are becoming so sophisticated that they are now used in clinical trials or medical research. These devices are made of sensors and they use this sensing capability to identify different activities or physiological information. Several types of physiological data is collected by wearable devices. Some of them are listed below.

1. **Heart rate:** Wearable devices like *smart watch*, *fitness trackers*, *ECG monitors* and *body mounted sensors* collect heart rate continuously. Some of the devices send notifications in case of any unusual response in heart rate.
2. **Respiratory rate:** *Smart eye wear* and *remote monitoring sensors* collect users respiratory rate.
3. **Activity:** There are different wearable devices that track a user's movements and activities, calculate active minutes and sedentary minutes. Devices like *smart band* (usually tracks fitness), *smart health watch* and *biosensors* collect physical activity and movement data.
4. **Glucose level:** A *glucose monitoring system* continuously collects the glucose level and notifies the user in case of any unwanted situation.
5. **Steps taken:** Different devices work as personalized systems and count the step taken by users and help them in maintaining their health.
6. **Blood pressure:** Blood pressure is collected by devices like *fitness trackers* and *remote monitoring systems*.
7. **Stress level:** Stress level data collected by different type of wearable devices such as *smart jewellery*.

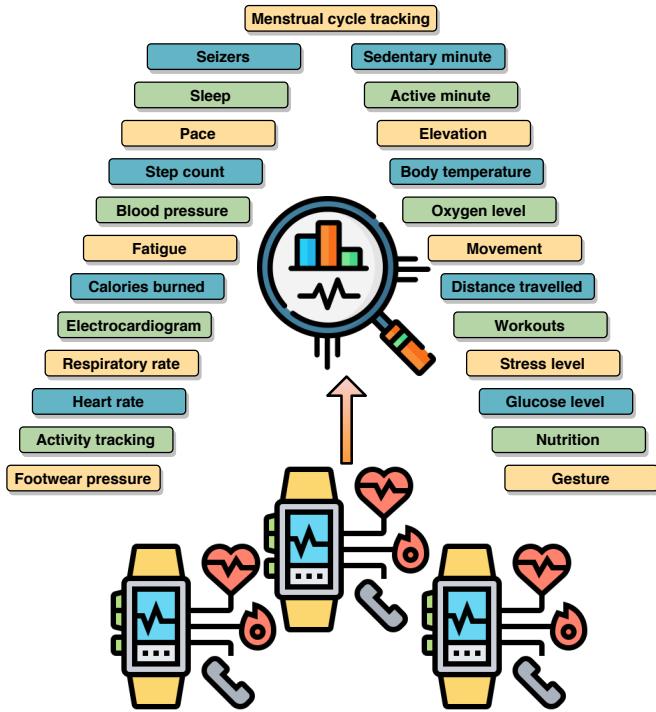


Fig. 7: Physiological data collected by wearable devices

8. **Distance travelled:** Devices like *fitness trackers* and *smart health watches* collect the data of distance and elevation.
9. **Calories burned:** *Smart bands* and *smart watches* collect the data of how much calories are burned and help the user in keeping track of their daily activities.

In addition to the above mentioned data, wearable devices also collect other physiological data such as oxygen level, menstrual cycle timing, body temperature and many more. Fig. 7 represents types of the physiological data generated and collected by different wearable devices.

### 3.3 Difference between Traditional Health Data and Wearable Health Data

Wearable data possesses some characteristics that differentiate them from traditional health data.

- **Continuous data:** Wearable devices continuously generate data. Different sensors continuously capture different physiological data. The devices capture data even when we are sleeping. On the other hand, traditional health data are generated mainly by the health professionals when we visit them.
- **Numerical Data:** generated by the wearable devices are mainly numerical, such as blood pressure measurement, heart rate and number of steps walked. On the contrary, traditional health record keeps record of our health con-

dition as diagnostic result or doctor's interpretation of the diagnostics.

- **Time-series data:** Wearable devices data are stored as the time series data where as traditional health data are stored as textual format in the database.
- **Real time data:** Wearable devices provide the opportunity to capture the physiological data in real time. On the other hand, traditional data are static and are not real-time.
- **Highly correlated:** Data points in wearable data are highly correlated.
- **More suitable for data analytic:** Wearable devices collect real-time data such as spatio-temporal data, trajectory data, location data but most importantly physiological data continuously (24/7) and in a format that is more consumable by the machine learning algorithms and can produce more accurate and effective data analytic.

## 4 Research

Through the SLR, we have investigated the existing research which have attempted to apply differential privacy in wearable data and tried to overcome the challenges identified in the previous section. A SLR is methodologically rigorous in contrast to ad-hoc reviews [33]. Our main focus is to identify relevant papers and review applications of differential privacy in wearable device generated physiological data, as well as to understand the conditions important for applying DP.

### 4.1 Research Questions

We have created five Research Questions (RQs), showed in Table 1 to guide our review.

Table 1: Research questions

ID	Research Questions
RQ1	What are the DP techniques that have been used in wearable data publishing?
RQ2	What are the major contributions of the proposed solutions in wearable data publishing?
RQ3	What types of datasets and programming languages are being considered for evaluation and implementation?
RQ4	What are the privacy criteria used in data publishing?
RQ5	What are the limitations of the proposed solutions?

## 4.2 Search Strategy

The overall search strategy is to find a body of relevant studies. Two search strategies, primary and secondary, have been used, as recommended by some studies [34, 35] to ensure that relevant studies have not been missed. In terms of record keeping, inclusion and exclusion strategies, we have followed PRISMA framework [36] (detail numbers are in the Appendix [Appendix A](#)). For the primary search, we have used search strings on several electronic databases. Following the primary screening, we have conducted a secondary search (paper selection) by means of backwards and forward tracing. The primary screening strategy involves search terms, literature resources and search process. These are described briefly as follows.

### 4.2.1 Search Terms

Throughout our searching process, we have considered journals and papers written in English. Besides this language factor, a date filter also has been applied. We have conducted our searches in several digital libraries. We have maintained a conceptual research string containing the main keywords of the theme. The search keywords are given in Table 2.

Table 2: Search terms / keywords

Number	Keywords
1	Review, survey, SLR, literature review
2	Wearable, medical, health data
3	Wearable devices generated data
4	Data publishing
5	Privacy preserving
6	Differential privacy
7	Temporal data

### 4.2.2 Literature Sources

In literature resources, we have conducted the searching process for papers on seven different electronic databases. During the paper collection process, we also have considered published journal names, published year, Computer Science Bibliographies, the title of the paper, the number of citations as well as the link of the paper.

After building conceptual search terms, we have used these keywords for finding journal papers, conference papers, and review papers in our considered electronic databases. Since different databases use different syntax for the search string, we have adjusted our search terms to accommodate with different databases. The search has been conducted on all the seven databases covering title, abstract, and keywords. The results of the search strings are given in Table 3.

Table 3: Number of papers retrieved from each digital library

Digital Library	No of Returned Papers
Google Scholar	14,435
IEEE	93
Springer	1,210
ACM DL	2,359
ScienceDirect	666
JAMIA	39
PubMed	17
<b>Total</b>	<b>18,819</b>

### 4.2.3 Search Process

SLR needs to comprehensively search all relevant sources; therefore, we have defined the search process by dividing it into the following two phases.

1. **Initial Searching Phase:** Searched in the seven electronic databases separately, and then gathered the returned papers together with those from a set of candidate papers. With the given search strings in Table 2, we have used appropriate logical operators (i.e., 'AND' and 'OR') along with parenthesis and quotation mark to refine our search and hence retrieved all the papers (details are in appendix [Appendix A](#)).
2. **Reference Searching Phase:** Scanned the reference lists of the relevant papers to find other relevant papers and then, if any, added them into the set.

We have used Microsoft Excel to store and manage the search results. We have gathered 18,819 papers from our initial searching phase and 13 papers from reference searching phase. Fig. 8 shows the search process in details including the number of papers.

## 4.3 Study Selection

We have found 18,819 candidate papers through Initial Searching Phase (see Fig. 8). Since many of our candidate papers will not provide useful information to address the research questions raised by this review, we have conducted further filtering to identify the relevant papers. More specifically, the study selection procedure has the following two phases:

1. **Initial Selection Phase:** We have applied the inclusion and exclusion criteria (defined below) to the candidate papers for filtering the relevant papers. These relevant papers can provide potential data for answering the RQs.
2. **Final Selection Phase:** In this phase, we have applied the quality assessment criteria (defined in Section 4.4) to the relevant papers for selecting the papers with acceptable quality, which are eventually used for data extraction.

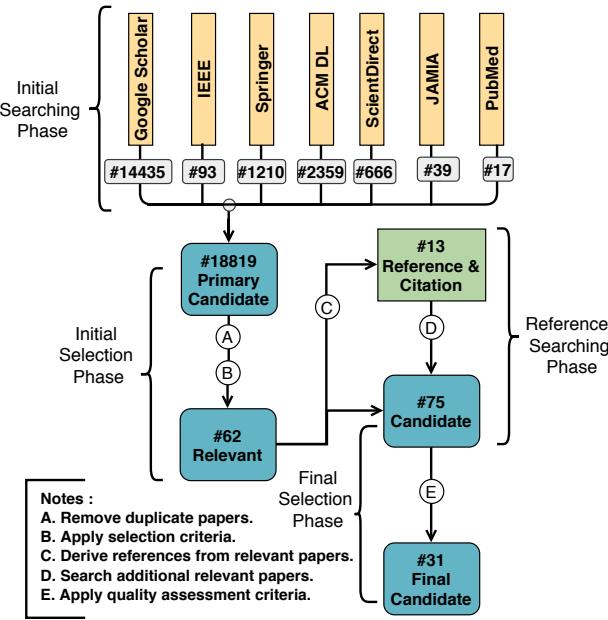


Fig. 8: Search and selection process

Our inclusion criteria are presented below:

- Abstract of papers written in English.
- Papers publish from 2007 onward.
- Papers publish until April 31, 2020.
- Academic papers published in conferences or journals.
- Papers that describe wearable or real-time data publishing other than trajectory or location data.
- For duplicate publications of the same study, only the most complete and newest one is included.
- For study that has both conference version and journal version, only the journal version is included.
- Review or Survey papers on real-time data publishing.
- Articles, Books related to wearable data publishing under differential privacy.

Next, the exclusion criteria are presented:

- Abstract of papers written in other languages.
- Duplicated papers found on the digital libraries.
- Papers worked on static or traditional health data.
- Editorials, prefaces, summaries, interviews, news, correspondences, discussions, comments, reader's letters, and summaries of tutorials, workshops, panels, and poster sessions.

We have extracted 62 papers through our inclusion and exclusion criteria. With that we have scanned the citation and references of these relevant papers and found 13 additional relevant papers which have been missed in the initial search process. Therefore, we have been able to identify 75 relevant papers. In the last step, we have applied a few quality assessment criteria for identifying the final selected studies, which are then used for data extraction.

Table 4: Quality assessment questions

ID	Quality Assessment Questions
QAQ1	Is the paper related to real-time or wearable data publishing under differential privacy?
QAQ2	Is the dataset used in the experiment real dataset or synthetic dataset?
QAQ3	Is the validation of the proposed method done using a dataset or not?
QAQ4	Does the study add value to a digital library of the industry community?
QAQ5	Are the limitations of study analyzed explicitly?
QAQ6	Is the proposed publishing method compared with other existing methods?

#### 4.4 Study Quality Assessment

We have created some Quality Assessment Questions (QAQs) in order to validate the quality of the papers. These questions are showed in Table 4.

The questions presented in 4 are used for the quality assessment of the research papers that we have collected. QAQ1 evaluates if the paper we are evaluating relates to wearable data publishing or real-time publishing or not. We have observed that researchers have evaluated their proposed DP-method using synthetic data rather than real wearable data; QAQ2 assesses this. If the dataset is a real wearable dataset, we keep them in our list otherwise, we discard them. In addition, some research works either do not have any validation or validation without a dataset; QAQ3 assesses that. QAQ4 evaluates if the research works adds any advance to the existing knowledge in terms of academic or industry practice. QAQ5 evaluates if the researchers have analyzed their own limitations or not. Finally, in QAQ6, we have checked whether the research paper has done any benchmarking with the existing work or not.

We have selected each paper based on the total number of QAQs they satisfy. We have read all 75 papers and evaluated them according to the Quality Assessment Questions. We have put a paper into final selected studies if the paper satisfies at least half of the QAQs. Finally, we have selected 31 papers. All the collected papers are listed in Table 5.

## 5 Analysis

Before jumping into analyzing the RQs, we have organized the papers into following three categories. All the papers in these categories have used differential privacy to preserve the privacy of the health data.

1. **Physiological** represents papers that have covered *wearable physiological data*. Physiological data include, but are not limited to, EEG, ECG, EMG, blood pressure, and

Table 5: List of selected papers

Category	Selected Papers
Physiological	Lin et al. [37], Lin et al. [38], Mohammad et al. [39], Prema et al. [40], Zhang et al. [41], Song et al. [42], Guan et al. [43], Kim et al. [44], Kim et al. [45], Lin et al. [46], Hao et al. [47], Kim et al. [48], Zhang et al. [49], Julian et al. [50], Zhang et al. [51], Nazir et al. [52], Bozkir et al. [53] Arijit et al. [54].
Real-time	Liyue et al. [55], Wang et al. [56], Rastogi et al. [57], Shi et al. [58], Yang et al. [59], Wang et al. [60], Gao et al. [61], Fan et al. [62], Kellaris et al. [63].
Others	Nguyen et al. [64], Yang et al. [65], Thomas et al. [66], Luo et al. [67].

activity. These papers have mainly discussed how to collect these types of data using wearables and publish them in a privacy preserving way.

2. **Real-time** represents papers that have discussed *real-time* and *dynamic* health data. This category only includes those papers which have discussed about real-time data collection or publishing. In terms of types of data, it is wearable health data. However, it is exclusively for real-time health data collection and publishing. So, if a paper discusses about historical physiological data then we have placed it in the previous (physiological) category, however, if it discusses real-time physiological data then we have placed it under this category.
3. **Others** represents papers that are not directly related to wearables, rather they are related to Medical Internet of Things (MIoT) and smart devices, such as mobile phone based healthcare.

### 5.1 RQ1: What are the DP techniques that have been used in wearable data publishing?

Wearable data has some characteristics which make them different from traditional health data, such as real time, dynamic, numerical and highly correlated data(details in 3.3). We have discussed different types of differential privacy techniques in 2.2. In this research question, we will explore which of these techniques and other DP techniques have been used by the researchers to protect the privacy of wearable data. Here, we have reviewed the proposed techniques and methods to publish wearable physiological data under differential privacy.

Applying DP is challenging for real-time or transaction data, because differential privacy was built for providing  $(\epsilon, \delta)$  privacy guarantee for statistical data only. Different researchers have addressed these issues and thus proposed different techniques for publishing real-time data while main-

taining differential privacy. In traditional differential privacy mechanisms, it is assumed that the data are independent, i.e. they are not correlated and the adversary does not have any knowledge of the data correlations. But real-time generated data can be correlated or we can acquire correlations among data.

Fig. 9 represents different types of DP techniques used for different types of wearable data (e.g., physiological, real-time, and others). It is evident from the figure that among all the techniques Laplace Distribution is the most popular for adding noise to the data. Other techniques are Geometric Distribution and Fourier Perturbation Algorithm (FPA). For eye data, researchers have preferred Gaussian noise for perturbation [50, 53]. For ensuring privacy guarantee of real-time and physiological data in health, techniques such as adaptive sampling [41, 55, 60, 62], filtering [41, 55, 60, 62], adaptive budget allocation [41, 60], filtering with Laplace distribution [60] have been used.

Researchers have extended Laplace distribution to provide better privacy guarantee. Shi et al. [58] used Symmetric Geometric Distribution (SGD) with Laplace distribution to provide discrete approximation to the Laplace distribution. Haar Wavelet technique [38], Bucket partition algorithm for partitioning dataset [39], Geometric technique [42] have also been adopted by different researchers. Despite the temporal correlation, researchers such as Rastogi et al. [57] and Bozkir et al. [53] have tried to achieve differential privacy using FPA.

In general, Laplace mechanism provides better accuracy compared to Gaussian mechanism. Therefore wearable data being mostly numerical, the Laplace mechanism may be an appropriate choice for perturbation.

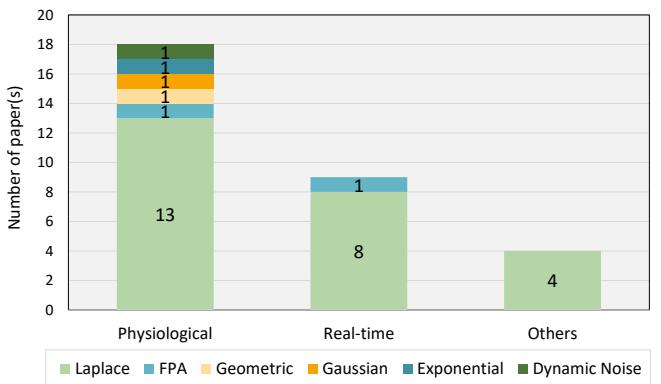


Fig. 9: Considered perturbation techniques for achieving differential privacy

Table 6: Research work related to physiological data publishing using DP

Paper	Name	Used technique (For RQ1)	Major Contribution (For RQ2)
Lin et al. [37]	-	DP-based Dynamic Noise threshold	A DP-based new scheme for large data from body sensor networks
Lin et al. [38]	-	DP (Laplace noise) with Haar Wavelet technique	Differentially private scheme for sensitive big data in BSNs with reduced errors.
Mohammad et al. [39]	-	Bucket algorithm and Laplace distribution	An efficient differentially private mechanism for releasing health data
Prema et al. [40]	-	Selective Gaussian mechanism with DP	A DP-based scheme for big data BAN which is more available and reliable
Zhang et al. [41]	RE-DPocpor	Laplace noise with adaptive sampling, filtering and budget allocation techniques	Dataset privacy where data that has been collected from $w$ -consecutive days
Song et al. [42]	PPM-HDA	Geometric Distribution	A more diverse and secure mechanism resisting differential attacks and supporting additive and non-additive aggregation
Guan et al. [43]	EDPDCS	K-means clustering and Laplace noise based DP	Proposed an efficient privacy-preserving clustering scheme over Map-reduce Framework for IoMT
Kim et al. [44]	-	Laplace distribution	Presented method is capable of preserving privacy of individuals sensitive data generated from smartwatches
Kim et al. [45]	-	Laplace distribution by leveraging LDP	Developed method can collect sensitive health lifelogs from smartwatch under DP
Lim et al. [46]	-	LDP with Laplace noise	Developed technology is capable of collecting health data from smartwatches by maintaining LDP
Hao et al. [47]	PMHA-DP	Laplace noise with a hierarchical method	Proposed multi-functional health data aggregation scheme under DP
Kim et al. [48]	-	LDP with Laplace	Proposed mechanism can collect individual temporal health data at fixed intervals by leveraging LDP
Zhang et al. [49]	WSV-MDAV	Laplace noise with micro aggregation algorithm	Proposed a privacy protection model based on aggregation algorithm for wearable devices using DP
Julian et al. [50]	-	Exponential Mechanism	Designed interface for VR to prevent user re-identification and protect gender information using DP
Zhang et al. [51]	APDP	Modified Laplace Mechanism	A fog computing based secured smart-home model with a personalized DP scheme
Arijit et al. [54]	-	Data driven technique & Laplacian noise for data obfuscation	On-demand obfuscation of sensitive data by satisfying DP
Nazir et al. [52]	mSieve	Laplace Distribution	Defined a new behavioral privacy metric under differential privacy
Bozkir et al. [53]	-	Fourier Perturbation Algorithm	A chunk based privacy-preserving method for eye movement features by considering different factors

## 5.2 RQ2: What are the major contributions of the proposed solutions in wearable data publishing?

In this section, we have explored the major contributions of DP-based techniques by different researchers in terms of reliability, utility, accuracy and risk minimization. We have divided the research works in three different categories.

### Physiological category:

In [37], authors have introduced a new DP-based concept called *Dynamic noise threshold* which is suitable for large amount of data from Body Sensor network (BSN). Similarly, in [38], authors have used differential privacy (Laplace noise) along with the *Haar wavelet technique* [68] for histogram to binary tree conversion to be used for sensi-

tive big data in BSNs. Using this approach, they have been able to reduce errors and provide long-range queries using a tree structure. Prema et al. [40] have introduced another DP-based Gaussian scheme for BSNs in which Gaussian noise is applied to only important features if DP does not produce a satisfactory protection. They have claimed their approach is *more reliable*.

In terms of improving the *utility*, Mohammad et al. [39] have proposed an efficient differentially private mechanism which adopts the bucket partition algorithm and Laplace distribution for preserving privacy. The effectiveness of their proposed technique is demonstrated by the overall improvement in the *accuracy* of perturbed data. In a similar domain, Zhang et al. [41] have adopted Laplace noise as the data

Table 7: Research work related to real-time data publishing using DP

Paper	Name	Used technique (For RQ1)	Major Contribution (For RQ2)
Rastogi et al. [57]	PASTE	Fourier Perturbation Algorithm (FPA) & Distributed Laplace Perturbation Algorithm (DLPA)	Combining FPA and DLPA to achieve the accuracy benefits of the former and the scalability of the latter
Shi et al. [58]	PSA	Symmetric Geometric Distribution (SGD) with Laplace distribution	Combining differential privacy and cryptography to calculate the approximate aggregate statistics for a time interval over encrypted data
Liyue et al. [55]	FAST	Laplace noise (using a white Gaussian error with variance)	Improved data accuracy using Kalman filter and privacy cost minimization using adaptive sampling
Fan et al. [62]	FAST	Laplace noise with filtering and adaptive sampling	Differential private real-time aggregate statistics based on filtering and adaptive sampling
Kellaris et al. [63]	BA, BD	Laplace noise with sampling and dynamic privacy budget allocation	Proposal of two novel mechanisms along with several optimizations
Wang et al. [56]	UKFDP	Laplace noise with unscented Kalman Filter	Kalman filter based DP for nonlinear systems enabling differentially private streaming data share
Yang et al. [59]	ConTPL	Laplace mechanism	A system to automatically convert an existing differentially private streaming data within a specific level
Wang et al. [60]	RescueDP	Laplace noise with adaptive sampling, budget allocation, dynamic grouping and filtering	Monitoring online aggregations of infinite streams with privacy guarantee
Gao et al. [61]	-	Laplace mechanism with GGA algorithm and Kullback Leibler (KL) divergence	Proposed approach can publish histogram for differentially private dynamic data based on Kullback-Leibler (KL) divergence

perturbation method along with adaptive sampling, filtering and budget allocation techniques. Their method allows the release of real-time health data with  $w$ -day differential privacy where the health data is collected for any consecutive  $w$  days. They have compared their technique to state-of-the-art methods for performance comparison and have proved that their method outperforms others in terms of *utility and privacy* guarantee.

To improve the *accuracy* of DP-based techniques, the authors in [44–46] have adopted the Laplace distribution noise as their base perturbation technique. According to [44], authors have asserted that their proposed approach can be used to *efficiently* compute population data while maintaining privacy through the use of a wristwatch. However, the

authors in [45, 46] have additionally leveraged LDP for collecting health data from smartwatches. Both the research works have conclusively contributed to successfully preserving the *usefulness* while properly gathering data from smartwatches with *privacy preservation* in place. Finally, Kim et al. [48] have proposed a novel mechanism to collect individual temporal health data at fixed intervals by leveraging (by using max advantage) Laplace Differential Privacy. As a result, their proposed technique have outperformed straightforward methods by delivering a significant improvement in *accuracy*.

Risk of privacy attack is major concern for any privacy preserving technique. Some researchers have worked to improve DP-based technique to *avoid the attack*. For instance,

Table 8: Research work related to others categories

Paper	Name	Used technique (For RQ1)	Major Contribution (For RQ2)
Nguyễn et al. [64]	Harmony	Local Differential Privacy (LDP)	An efficient solution for smart device data using LDP
Yang et al. [65]	MLDP	Machine learning with Laplace noise (noise added in training set)	A ML based differentially private aggregation method in IoT within a fog computing architecture to reduce communication overhead and release cloud burdens
Thomas et al. [66]	-	Laplace distribution and Sine polyonym	Utility maximization with adjustable privacy settings for calculating aggregations over private sensor data
Luo et al. [67]	Salus, P3	Dynamic Noise by leveraging Laplace Distribution	An input perturbation algorithm to preserve DP by providing strong resilience against data reconstruction attacks and predictable utilities

an aggregation scheme named *PPM-HDA* have been proposed in [42] which supports both multi-functional additive (average, variance) and non-additive aggregation (min/max, median, sigma-percentile, and histogram). It is claimed that the proposed mechanism is more diverse and secure for cloud servers, *resisting differential attacks*. Authors in [43] have proposed a clustering scheme which introduces a privacy-preserving clustering scheme for the Map-reduce framework with *improved accuracy* by optimizing privacy budgets. To achieve this, they have used K-means clustering and Laplace noise for DP. Zhang et al. [51] have proposed a fog computing based smart-home model and explored collision attacks under personalized protection scenarios using DP. In their proposed model, noise is generated under a Markov process and the *privacy protection* is achieved using a modified Laplace distribution. Their experiment have resulted in successful *privacy enhancement* while minimizing overall privacy budget and *eliminating background knowledge attack*. Authors in [49] have identified and solved the issues of V-MDAV algorithm and then, have proposed a privacy protection model, based on an aggregation algorithm, named *WSV-MDAV* for wearable devices using DP. According to their experimental assessment, their technique have *improved privacy protection performance* and *reduced data loss* when compared to the traditional method. In [54], Arrijit et al. have proposed a solution that can obfuscate any sensitive data on-demand by satisfying differential privacy. This work is of practical importance that have *improved the performance* by minimizing the privacy breaching risk.

*Reduction of computational overhead* is also an active research direction. In [50], authors have designed a Virtual Reality (VR) interface which *prevents user re-identification* as well as *protects gender information* by using DP. Their experiment is effective in *reducing overhead*, resulting in a *low-cost solution* for preserving users' privacy while preserving utility. In [53], Bozkir et al. have put forward a chunk based privacy-preserving method for eye movement features by considering the factors reduction of query *sensitivity, complexity* and *temporal correlations*. Their transform coding based solution is claimed to be *more adaptive* than various existing low-complexity methods. Both these papers are related to eye movement data.

A summary of the used techniques and major contributions in the research papers under the physiological category is presented in Table 6.

#### Real-time category:

In [57], authors have proposed a scheme for real-time health data named PASTE where both the Fourier perturbation algorithm and Laplace distribution are used. By perturbing Discrete Fourier Transform (DFT) of query answers, the proposed FPA algorithm can answer multiple queries over time-series data and ensure DP despite the presence of a temporal correlation. On the other hand, the proposed

DLPA (Distributed Laplace perturbation Algorithm) can be used for adding noise in a distributed way, a useful feature in the absence of a trusted third party. By combining FPA and DLPA, PASTE gets the *accuracy benefits* of the former and the *scalability* of the latter. In [58], authors have proposed a solution by combining differential privacy and cryptography enabling a user to upload a stream of encrypted data to an aggregator (can be untrusted) and the aggregator can calculate the approximate aggregate statistics for a time interval through the proposed algorithm. Combining these methods have helped them achieving *strong privacy guarantee*.

Some researchers have used adaptive sampling and filtering to preserve the privacy of real-time data to improve *utility and performance*. For example, in [55], the proposed approach enables to release time-series data under differential privacy by *improving data accuracy* (using Kalman filter [69]) and *minimizing overall privacy cost* (adaptive sampling algorithm with PID control). Similarly, Fan et al. [62] have proposed a framework to release real-time aggregate statistics by satisfying differential privacy based on filtering and adaptive sampling. Their adaptive methods *improves the utility* and demonstrates excellent *performance* even under small privacy cost.

In [63], authors have considered sliding window methodology using Laplace noise with sophisticated sampling and dynamic privacy budget allocation techniques. This will improve the *scalability* in terms of real time data publishing. They have also proposed three benchmark methods named *FAST<sub>w</sub>*, *Uniform*, and *Sample*. The solution is based on Kalman filter based differential privacy to facilitate streaming data sharing. It takes advantages of sigma points [56] for non-linear systems. This method *increases the accuracy* of the published data [55].

To *overcome the temporal correlation* problem in real-time data, Yang et al. [59] have designed system that can automatically convert an existing differentially private streaming data into one bounding Temporal Privacy Leakage (TPL). On demand sensitive data obfuscation is also used for real-time streaming data. Furthermore, authors in [60] have proposed a framework named RescueDP which can monitor the online aggregation of an infinite stream with privacy guarantee. Using adaptive and dynamic methods RescueDP outperforms existing methods and *preserves utility* with proper *privacy guarantee*.

Finally, the proposed algorithm by Gao et al. [61] can publish histograms for differentially private dynamic data based on Kullback-Leibler (KL) divergence [70]. Using this methods have resulted in overall *accuracy improvement* and *utility enhancement*.

A summary of the used techniques and major contributions in the research papers under the real-time category is presented in Table 7.

**Others category:** Thomas et al. [66] have proposed to select the proper privacy settings to calculate an aggregation function over private sensor data. It can also help to *maximize the utility* for different level of privacy. This makes the method *secure and reliable*. In [65], authors have considered a fog architecture instead of cloud architecture, where their proposed multi-functional aggregation method *reduces communication overheads* as well as *releases cloud burdens*. Considering data reconstruction attacks, Luo et al. [67] proposed an input perturbation algorithm *Salus*. This light-weight algorithm provides *strong resilience* against data reconstruction attacks while preserving differential privacy. Later *Salus* was extended in P3 framework for supporting privacy-preserved Mobile Crowdensing Services (MCS) applications [65]. Authors in [64] have proposed a system that is *practical, accurate, and efficient* for gathering and examine data from smart device users under LDP.

If we review the key contributions of researchers to differentially private health data publication discussed in this article, we can demonstrate that the most important aspect that have been considered by researchers are *privacy and utility enhancement*. Due to the fact that differential privacy involves a trade-off between privacy and utility, it is critical to address privacy preservation in a way that does not jeopardize the utility of data. Additionally, researchers have concentrated on improving the *overall performance* of their proposed mechanism in order to outperform existing works. Improving mechanism *accuracy* have also been a center of focus of the researchers. Additionally, the researchers have focused on *overhead reduction* and *secure and reliable* model building. Although It has been noticed that researchers have placed a greater emphasis on utility, privacy and performance enhancement compared to the model's security and reliability.

A summary of the used techniques and major contributions in the research papers under the physiological category is presented in Table 8.

### 5.3 RQ3: What types of datasets and programming languages are being considered for evaluation and implementation?

Validating any research proposal is a very important part of the research. Researchers employ different types of methods to validate their proposed methods, systems, protocols, or techniques. In a data-driven scenario, the quality of validation heavily depends on the quality of the dataset. In this section, we have reviewed the types of dataset the researcher have considered to evaluate their research proposal. We have divided the dataset into two categories based on their availability, such as

#### 1. Public datasets: Datasets that are publicly available.

#### 2. Private datasets: Datasets that are self made or self collected by the researcher. We have considered synthetic datasets also a private datasets.

**Physiological category:** In the physiological category, a large variation of datasets have been utilised. In terms of heart related dataset, Zhang et al. [41] have collected heart rates data for three months from hospital patient. Nazir et al. [52] have also used a private dataset containing 660 hours of ECG data collected from 43 participants. Lin et al. [37] used a private dataset from wearable sensors where they have collected ECG data of 2.2 millions data points. Heart disease dataset is also used in the research [40]. Mohammad et al. [39] have generated a private dataset of heart rates from wearable devices which were attached to a user for two weeks. On the other hand Guan et al. [43] have utilised blood record dataset in their research. The blood dataset contains individual information of blood donation and the other dataset contains the identity of the individuals and other general information.

In addition, many researchers have worked with activity type data such as walking, running, sleeping. Kim et al. [44] have prepared a dataset for daily step counts collected using Gear S3 smartwatch between a limited time period and then replicated 10, 100 & 1000 times. In another work, Kim et al. [45] have utilised a private data set consisting of daily cumulative step-count data of 247 days where each cumulative step-count data corresponds a stream of length 600. Kim et al. [48] have also used a public PAMAP2 physical activity monitoring data collected from [71] which contains a heart rate monitoring dataset that is collected using sensors. To track physiological activity, smart-home data is also used [72, 73]. Data are collected under 7 scenarios, including sleeping, resting, dressing, eating, toilet use, hygiene and communication.

Bozkir et al. [53] also used two public datasets: i) public eye-tracking dataset collected with an Oculus VR device and ii) pupil eye-tracking dataset from [74] where 20 participants were tasked with reading three different document types (a comic, newspaper, and textbook) in a VR environment. The utilisation of different datasets is summarised in Table 9.

**Real-time category:** Real-time data publishing is also validated by using datasets. Liyue et al. [55] have utilised three different sets of public datasets, namely Flu dataset [76], Traffic dataset [77] and Unemployment dataset [78] and tried to correlate them. The same Flu dataset has been harnessed by Gao et al. [61] as well. Rastogi et al. [57] have utilised different public datasets such as GPS, Traffic and Weight where the last dataset contained daily weight data of about 300 users. Fan et al. [62] have utilised three real-world public datasets which are Flu [76], Unemployment [78] and Traffic [79]. Wang et al. [60] experimented with two real-world public datasets, Taxi Trajectory Prediction [80] and

Table 9: Different types of datasets

Data Type	Availability	Physiological	Real-time	Others
Heart-related	Private	[37], [39], [41], [52]	-	-
	Public	[40], [54]	-	-
Blood dataset	Public	[75]	-	-
Activity (Step Count, Running)	Private	[44], [45]	-	-
	Public	[48]	-	-
Eye tracking	Public	[53]	-	-
Mix	Wearable sensors	[38], [37], [46]	-	-
	Smart home	[51]	[56]	-
	GPS, Traffic, Weight	-	[57], [60]	-
	Flu, Traffic and Unemployment	-	[62], [55], [61]	-
	Microsoft Band and Community Health	-	-	[67]
	Mobile health	-	-	[65]

World Cup [81] and one synthetic Spatio-temporal dataset called Brinkhoff [82]. Utilised datasets for the real-time category are summarized in Table 9.

**Others category:** Yang et al. [65] have used two real-world public datasets, namely Reference Energy Disaggregation Dataset and Mobile Health Dataset which consists of 1 million records from 24 different sensor signals. Luo et al. [67] also have used two real-world private case studies. The first one was a community health survey consisting of the heart rate of 20 students in order to find the average heart rate and heart rate distribution. The second dataset was a collaborative emotion classification dataset in which Microsoft Band was used to collect the heart rate, GSR, and skin temperature from users to build collaborative classification models. We have summarised the used datasets in the others category in Table 9.

**Programming languages:** In addition to dataset, we have also reviewed what programming languages have been used to implement differential privacy. In our survey, we have found that *Java*, *MATLAB*, and *Python* are frequently used by researchers for implementing their algorithms. Fig. 10 shows how different programming languages are used for implementation. However, recently python has emerged as the most used programming language in terms of open source differential privacy implementation [83]. Google has also open sourced their differential privacy library [84], which is implemented using C++ language.

#### 5.4 RQ4: What are the privacy criteria used in data publishing?

Different researchers have considered different privacy criteria for preserving privacy of data. Most common criteria is  $\epsilon$ -differential privacy( $\epsilon$ -DP) where  $\epsilon$  is the privacy budget associated with any data release. Other than this, for

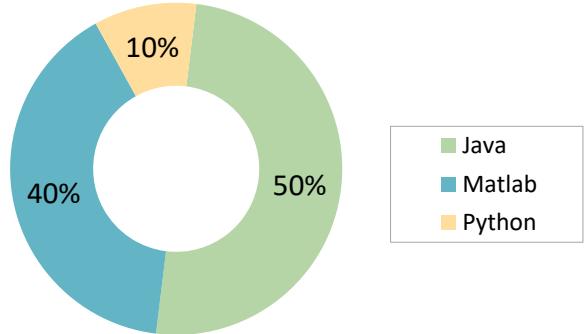


Fig. 10: Languages used for perturbation schemes

real-time data, Shi et al [58] have provided a methodology that depends on  $(\epsilon, \delta)$ -differential privacy( $(\epsilon, \delta)$ -DP). At the same time some have considered new privacy criterion such as w-event and w-day  $\epsilon$ -differential privacy [56, 60, 63]. In these mechanisms, privacy budget is calculated for w consecutive events. Liyue et al. [55] have solved the privacy preserving problem a bit differently. Their proposed solution satisfies  $\alpha/T$ -differential privacy, where  $\alpha$  is the privacy budget and  $T$  is the length of the entire series. Wang et al. [56] have collected health data for w different days and preserved privacy for those consecutive days. We have presented a summary of different privacy criteria in Table 10.

#### 5.5 RQ5: What are the limitations of the proposed solutions?

Researchers have conducted several experiments for evaluating their proposed methodology. However, these experiments have several limitations. The limitations of existing methodologies need to be addressed for a suitable, and more practical privacy preserved framework.

Table 10: Different privacy criterion adopted by researchers

Category	Privacy Criterion	List of Papers
Physiological	$\epsilon$ -DP	[40], [39], [51], [49], [42], [43], [44], [45], [48], [46], [47], [52], [53], [54]
	$w$ -day $\epsilon$ -DP	[41]
Real-time	$\epsilon$ -DP	[57], [59], [61]
	$(\epsilon, \delta)$ -DP	[58]
	$w$ -event $\epsilon$ -DP	[56], [60], [63]
	$(\alpha/T)$ -DP	[55]
Others	$\alpha$ -DP	[62]
	$\epsilon$ -DP	[64], [65]

**Physiological category:** For physiological data, the proposed DP solutions mainly suffer from scalability issues [52]. Many of the proposed differential privacy models are strict to static data publishing and confined to single dimension [49]. Moreover, many of the privacy protection schemes are just theoretical in nature [38, 50]. Some models [51] suffer from performance degradation with an increasing number of cloud resources. The method in [42] introduced relative errors with a heavy computational burden on cloud servers. Finally, the algorithm proposed by Mohammad et al. [39] has a higher complexity than existing works (e.g. Li's [85]). In addition, the proposed methods also vulnerable to information leakage in the presence of a strong adversarial model. It can cause the adversary gain more knowledge and result in privacy leakage.

**Real-time category:** For real-time data, the proposed DP solutions mainly suffer from different types of errors such as reconstruction & perturbation errors [57], relative errors [56, 62] and absolute error [61] have been observed. To detail, an algorithm's accuracy is found to be affected by failures occurred when the algorithm is executing to answer a query [57]. For non-linear synthetic datasets, the proposed method has a higher relative error due to a model misfit compared to the existing methods [62]. One of the most common problem is the difficulty choosing an optimal value for epsilon ( $\epsilon$ ) to gain any advantage. For example, for a large budget ( $\epsilon > 1$ ), there is no substantial advantages [55] and even some proposed method under perform with a higher epsilon value [56]. Similarly, in [61], it has been found that increasing in epsilon value has weakened the algorithm. Thus, choosing an appropriate value for threshold is a challenging task. Besides, the absolute error in the GGA algorithms is smaller than other algorithms only when the query range is greater than 40. Authors in [58], have noted a number of issues such as dynamic join and aggregation problem. Also, when a node failure occurs, some of the participants become unable to provide their encrypted data. Another limitation is

lack of support for graceful degradation during node failures.

**Others category:** As for the others category, communication overheads with estimation errors have been reported as the major shortcoming [64]. In [66], local and global errors have been highlighted, compared to the Laplace mechanism, because of the usage of sine polyonym. System and computational overheads have incurred due to usage of Salus in [67]. In addition, computational overhead and data reconstruction error have been experienced. Maintaining a balanced noise and sensitivity have also been an obstacle in the way of preserving privacy as large training sets contain too much noise which results in the loss of utility for the proposed model [65].

**Summary:** Different types of limitations found in different approaches are summarized in Table 11.

## 6 Discussion

Differential privacy has paved the way for a more flexible solution in privacy preservation. It has overcome the limitations of existing methodologies to some extent. However, the basic differentially private perturbation method alone cannot protect data from getting exposed. Several researchers have identified the major concerns regarding challenges of publishing such data by using basic mechanisms of differential privacy [86, 87]. Therefore, researchers have proposed to combine different methods with differential privacy in order to provide an effective privacy protection mechanism. They have tried to propose their mechanisms in such a way that both the privacy and data utilization are well balanced.

Throughout the paper we have explored, through our research questions, the advantages of applying differential privacy over other techniques, challenges faced by basic mechanisms and how different researchers have extended the basic differential privacy to meet the requirements of wearable data publishing. In this section, we have summarised our findings from different perspectives.

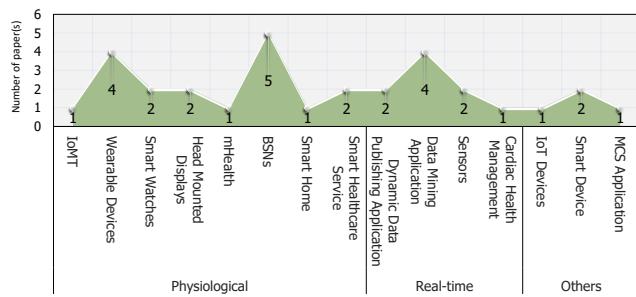
### 6.1 Research Question Perspective

At first, we present the following summary from the perspective of our research questions.

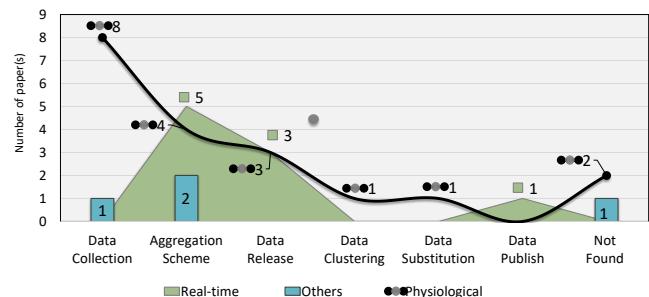
- **Application domains:** Researchers have considered different application domains for their proposed mechanisms. We have illustrated the number of published papers for different domains in Fig. 11a. As evident from the figure, most of the researches have explored areas such as BSNs, wearable devices, data mining applications and so on.

Table 11: Limitations in the current research works

Limitations	Type	Physiological	Real-time	Others
Error	Relative error	-	[62], [42]	-
	Reconstruction error	-	[57]	-
	Perturbation error	-	[57]	-
	Absolute error	-	[61]	-
	Error increased due to larger group	-	[63]	-
	Estimation error	-	-	[64]
	Data reconstruction error	-	-	[67]
Overhead	Communication Overhead	-	-	[64]
	Computational overhead	[42]	-	[67]
	System Overhead	-	-	[67]
Appropriate value for epsilon	-	[38]	[55], [56], [61], [62], [63]	[65]
Algorithm	Complexity	[39]	-	-
	Unable to detect tempered data	[47]	-	-
	Scalability	[52]	-	-



(a) Different application domains



(b) Different data management techniques

Fig. 11: Number of research papers in different application domains and data management

– **Application scheme:** Versatile application schemes have been considered by researchers while preserving data privacy. Some researchers have considered data publishing scheme whereas some of them have proposed aggregation or data release schemes. Fig. 11b visualizes the proposed schemes. From the figure, for the physiological category, the highest number of papers (8) are for the data collection scheme. On the other hand, the aggregation scheme has the highest number of papers, with 5 and 2, for the real-time and others category respectively.

– **Privacy criteria:** In case of privacy criteria, researchers have mostly considered  $\epsilon$ -DP. Other than this, they have also considered  $(\epsilon, \delta)$ -DP,  $\alpha/T$ -DP. Some of them even considered  $w$ -event DP where data is collected for consecutive  $w$ -events. Fig. 12 illustrates privacy criteria of proposed schemes in each category: physiological, real-time and others.

– **Programming language used:** For implementing differential privacy algorithms, different programming lan-

guages such as *Java*, *Python*, *Matlab* have been considered. We have presented a detailed summary of programming languages used in developing the mechanisms in Fig. 13.

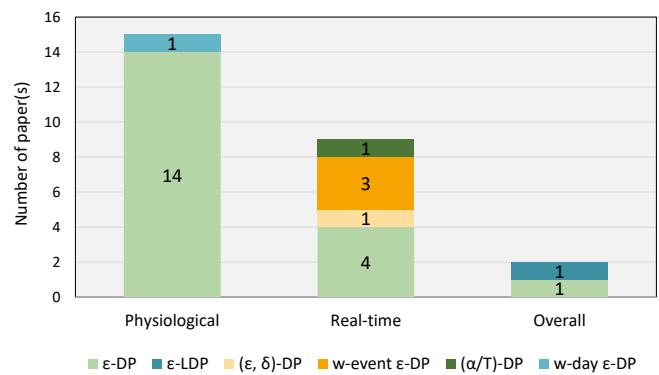


Fig. 12: Privacy criteria of proposed schemes

**– Considered datasets for developing mechanisms:** After developing the mechanism the most important thing to do is to evaluate the mechanism. For this purpose researchers have used both the private and public datasets. We have provided a summary in Fig. 14 which visualizes datasets used by different researchers.

## 6.2 Data Perspective

Next, we have summarised all works from the perspective of data: physiological, real-time and others.

**– Physiological data:** With the enormous advancement in sensor technology and the physiological data generated from them, the privacy issues with physiological data is a concerned topic. These sensors generate data which can be dynamically updated or can be temporally correlated. Although, *encryption* and *k-anonymity* are largely adopted solutions for preserving privacy, however, they can be more computationally complex and less practical. Conversely, differential privacy based solutions are more light-weight, more practical and thus have less communication overhead.

18 papers have been selected which have met our selection criteria for the physiological data category. In these 18 papers, researchers have showed various methods for preserving privacy in a better way. Among them, different application behavior such as data collecting, data releasing, aggregation studies for making decisions are majorly noticed. These frameworks have covered areas such as BSNs, wearable devices, smart watches and even in head mounted displays. We have also noticed researchers used Laplace perturbation method most. Which means, numerical queries have been majorly covered. Throughout the method, mostly they have followed  $\epsilon$ -DP. Other than this, FPA, Gaussian mechanism and  $w$ -event privacy have also been applied. For conducting experiments, in most cases, they have developed their algorithm in Java and considered their self generated dataset.

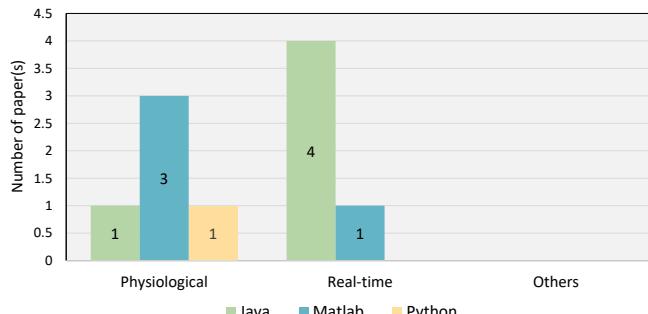


Fig. 13: Languages used for developing mechanisms

Fig. 15 makes a summary what we have found among the selected 18 papers.

Researchers have focused on various aspects of physiological data. They have focused mainly on effectiveness of publishing data, how efficiently data can be collected, reducing calculation and communication overhead, maintaining availability and reliability so that proposed solutions provide better accuracy as well as maintain the balance between utility and privacy. Differential privacy based schemes such as, MHDA<sup>⊕</sup> [42], ReDPoctor [40], EDPDCS [43], WSV-MDAV [49], APDP [51] outperforms other existing methods. PMHA-DP [47] has less communication overhead than existing solutions. APDP gives more privacy protection in comparison with UDP and NPDP. It also has best performance in terms on attack resistance. The proposed method from [53] is capable of handling correlation in data.

**– Real-time data:** Differential privacy has also proven itself as a suitable solution in case of preserving privacy of real-time data. We have conducted our review over real-time healthcare sectors as real-time data also generates continuous data. We have found 9 papers satisfying our selection criteria, where majority of researchers have focused in aggregation studies by which data analytics can take important decisions. Other than aggregation, data releasing and publications are barely covered. These frameworks have covered areas such as data mining applications, sensors and dynamic applications as well.

From our review, we can conclude that the most of the researchers have considered Laplace mechanism as a mean of preserving privacy. Researchers have maintained  $\epsilon$ -DP for their given mechanisms. Besides, newly adopted privacy criteria such as  $w$ -event DP,  $\alpha/T$ -DP has also been observed. In general, developing mechanism in Java and conducting experiment with public dataset is preferred by the researchers.

For real-time health data, differential privacy based mechanisms outperform existing methods. Researchers have found better results by using differential privacy as a means of privacy protection. Rastogi et al. [57], Liyue et al. [55], and Fan et al. [62] have proposed methods that can achieve better accuracy and utility. The proposal from Rastogi et al. [57], PASTE, can also perform excellently under small privacy cost. RescueDP [60], an aggregate monitoring scheme, outperforms existing solutions as well as improves utility, ensures strong privacy guarantee. Proposed solutions of Wang et al. [56] using Laplace noise with unscented Kalman Filter and Kellaris et al. [63] using Laplace noise with sophisticated sampling and dynamic privacy budget allocation technique are more practical.

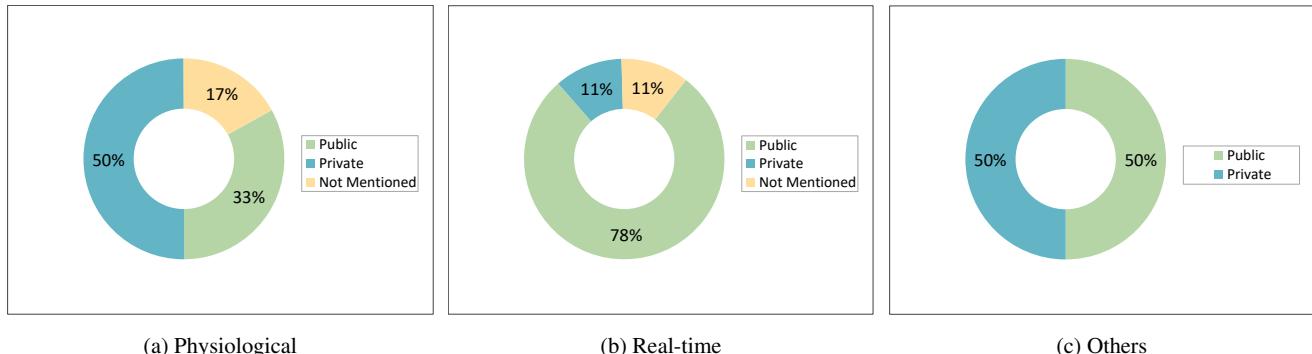


Fig. 14: Type of datasets used for evaluation

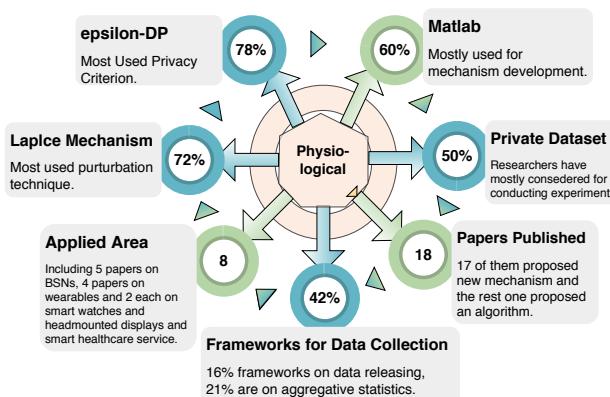


Fig. 15: Summary of the reviewed works in the physiological category

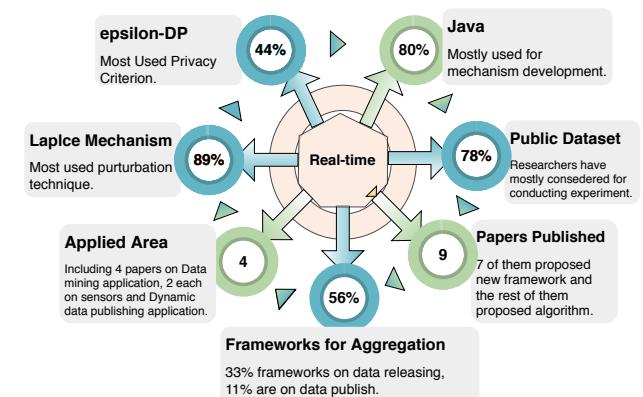


Fig. 16: Summary of the reviewed works in real-time category

Other mechanisms such as UKFDP [56], FPA<sub>k</sub> [57], ConTPL [59] and FAST [62] outperform the existing state-of-the-art methods and provide more practical solutions in terms of accuracy and privacy. Gao et al. [61] have provided a solution that can reduce noise errors and outperform existing solutions.

In Fig. 16 we have represented a summary for real-time healthcare domain.

**- Others:** The papers in the others category generate similar patterns like wearable or real-time health data. Differential privacy has also been found beneficial for applying privacy-preserving mechanisms over these data. We have reviewed 4 papers which have satisfied our selection criteria.

The frameworks proposed in these papers cover a wide range of application areas such as smart devices, IoT devices and MCS application. Among all the mechanisms, Laplace mechanism has been mostly adopted by researchers with  $\epsilon$ -DP being the most widely used privacy criterion. But none of the researchers has mentioned anything about the platform/language they have used for

developing either the framework or algorithm. The researchers have conducted various experiments in order to prove the efficiency of their proposed solution and they have used both the public and private datasets equally for their experiments. Among the solutions, Harmony [64] and Salus [67] are more practical, provide accurate and efficient results, reduce errors and maintain a stable balance between privacy and utility by trying to improving accuracy. With the help of fog computing architecture, MLDP [65] can aggregate by reducing communication overhead and [64] can reduce both the computational and communication overhead. Researchers of [67] have claimed that they achieved enhancement in data protection by using differential privacy. In Fig. 17, we have represented a summary for the others category.

## 7 Current Challenges and Future Direction

Table 12 provides a summary of the major contribution in terms of wearable data publishing under differential privacy. We have observed that researchers have focused more on

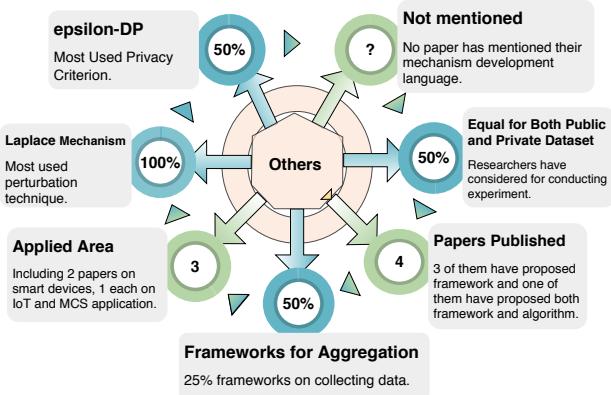


Fig. 17: Summary of the reviewed works in the others category

Table 12: Major contributed area by research communities

Contribution Type	Physiological	Real-time	Others
Privacy Enhancement	[40], [37], [49], [51], [53], [54]	[60]	[65], [67]
Utility Enhancement	[51], [52], [45], [46]	[56], [60], [61], [62]	[64], [67], [66]
Improved Accuracy	[38], [39], [43], [45], [46], [52]	[61]	[64], [65]
Performance Enhancement	[41], [48], [43]	[55], [60], [63], [56], [63], [57], [62]	[64], [67]
Reduction of Overhead	[47], [42], [38], [50]	[55]	[64], [65]
Secure and Reliable	[42], [51]	[56]	[67]

utility and performance enhancement compared to security and reliability of the model.

### 7.1 Limitation of the Existing Studies

We have compiled the limitations of the current approaches and future research directions in the following section.

- Choosing appropriate value for  $\epsilon$ :** One of the major concerns in differential privacy is to choose an appropriate value for privacy budget denoted with  $\epsilon$ . The value of  $\epsilon$  determines the strictness and strength of privacy. A smaller value of  $\epsilon$  provides stronger privacy, however, with that the data losses its utility and vice versa. Therefore, finding a optimal value for  $\epsilon$  is a great challenge for any DP-based technique. There is very limited works have been done on finding the optimal value.
- Correlation of data:** Real-world datasets often contain strong correlation among the data which can cause disclosure of an individual's information. For example, such

correlation between data can enable an adversary to find out sensitive information about different individuals. The adversary can combine obfuscated data with existing correlation and derive sensitive information about individuals. Researchers have proposed model based approach [88], [89] and transformation based approach [90], [57] for solving this issue of data correlation. However, these approaches did not prove to be an optimal solution and even sometimes can distort the data to a great extent [91]. Therefore, overcoming the obstacles of data correlation is a big challenge for differential privacy.

- Sensitivity:** The principal purpose of differential privacy is to maintain the indistinguishability between the presence or absence of any individual in the dataset. Sensitivity is the maximum difference between two neighboring dataset (datasets differing in one row). Noise is added to cover the difference and maintain the same identity for both the databases. To improve sensitivity more noise needs to be added. However, large value of noise can distort the data and this can result in unwanted utility loss. These trade-off between privacy and utility needs to be maintained. Some technologies are using diversity sensitivity to overcome this issue [92]. However, it is still a challenge to choose an optimal value of sensitivity and preserve both privacy and utility simultaneously to maintain the trade-off.
- Vulnerability of basic mechanisms:** Basic mechanisms of differential privacy face various challenges when researchers tried to implement them. In [93], authors have shown Laplace Noise is vulnerable to tracker attack. After querying few times, results (after adding Laplace noise with true value) have either no privacy or no utility. In addition, [67] have shown Laplace mechanism is vulnerable to Data Reconstruction attack.

Finally, we have analyzed the papers published until April 31, 2020. Between May to the acceptance time, new papers may have been published in different scientific journals. In addition, we have not considered papers that are not focused on wearable devices rather used traditional IoT devices or MiIoT devices.

### 7.2 Future Direction

In this section, we discuss future research direction in differential privacy.

**Adaptive Privacy Budgeting for real-time data:** Selecting an appropriate  $\epsilon$  value is a crucial task in order to protect the privacy of the individuals. Unlike static data, we do not have prior knowledge of the data point values for the real-time streaming data. Thus, distributing budget adaptively (rather than statically) could be an excellent way to preserve balance between privacy and utility. Kellaris et al's [63] re-

search work has established the superiority of adaptive budget allocation over static budget allocation for streaming data.

**Integrating Blockchain with Differential Privacy:** In the past few years, blockchain has emerged as a key technology that establishes trust among trust-less parties in a distributed and decentralized manner. It has the potential to transform the way in which we share information [94] and guarantees secure and immutable data storage. Along with its association with the bitcoin concept, the blockchain has been widely adopted in various fields, including healthcare, finances, logistics, IoT [95–98], and even wearable devices and smart healthcare [99]. However, privacy is a big concern for blockchain, specially for public blockchain. Several researchers have used differential privacy to overcome privacy issues in blockchain systems [100].

In addition, blockchain is also used to build trust on the privacy budget by providing a distributed and transparent system. Authors from [101, 102] have proposed a blockchain based approach for tracking and saving differential privacy costs.

**Differential Privacy in Big data and Artificial Intelligence (AI):** In today's world big data and AI have become one of the major driving forces. In recent years, big data and artificial intelligence (AI) have gotten a lot of attention and have become valuable resources. Big data refers to the production of a massive amount of data from various sources, including sensors, wearable devices, IoT devices, social media platforms, and many more. Due to its massive scale, privacy and security are a major concern regarding this. Researchers are using differential privacy for big data publishing in different domains (e.g., transport, health)[103]. Authors from [104] have shown that integrating differential privacy have resulted in resolving many of the privacy issues of big data publishing.

## 8 Conclusion

Due to the rapid growth of wearable technologies, there is hardly any area which is not affected by it. Therefore, Health sectors are benefited by adapting wearable technologies. However, privacy is always a big concern for health data. Differential privacy has emerged as one of the most popular privacy preserving mechanisms in recent times. The purpose of this article is to understand the trends and limitations of differential privacy on wearable data.

Even though the existing privacy-preserving mechanisms are applicable in wearable technologies, there is still a gap, which necessitates additional effort in designing and developing a more secure privacy preserving mechanism to hinder PII information. Though the proposed schemes from [42, 64, 67] have outperformed existing state-of-the-art solutions by

providing a more efficient solution and have provided better data protection still these schemes suffer from overhead problems such as computational, communications, and even in-system overheads.

There are still a number of issues faced by differential privacy such as fine-tuning  $\epsilon$ , its privacy budget, to balance between privacy and usability as well as other issues such as dimensionality and temporal correlation also need to be addressed. There are some open research problems such as data reconstruction errors [57, 67], perturbation errors [57], absolute errors [61] and relative errors [42, 56, 62]. It is important to find solutions that can minimize an error rate significantly so that data utilization can be increased. In addition, we have also observed that there are only limited number of works have been done on real-time health data publishing. Finally, privacy mechanisms need to be more adaptive where users can fine-tune their privacy according to their needs.

## References

- Armonk. Ibm and partners to transform personal health with watson and open cloud. <https://www-03.ibm.com/press/us/en/pressrelease/46580.wss> (13 Apr 2015). [Online; accessed August 17, 2020]
- M. Gowtham, S.S. Ahila, in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)* (IEEE, 2017), pp. 1–5
- B. Yüksel, A. Küpcü, Ö. Özkasap, Future Generation Computer Systems **68**, 1 (2017)
- H.C. Assistance, Office for Civil Rights (2003)
- C.M. O'Keefe, C. Connolly, Electronic Journal of Health Informatics **6**(2), 16 (2011)
- Grubb, Ben. Thousands of medical histories exposed in data breach. <https://www.smh.com.au/business/companies/thousands-of-medical-histories-exposed-in-data-breach-20190807-p52euq.html> (August 7, 2019). [Online; accessed 7-November-2020]
- D.K. Altop, A. Levi, V. Tuzcu, in *2015 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)* (IEEE, 2015), pp. 92–99
- M.J.M. Chowdhury, T. Pal, in *2009 International Conference on Electronic Computer Technology* (IEEE, 2009), pp. 541–544
- Y. Ji, J. Zhang, J. Ma, C. Yang, X. Yao, Journal of medical systems **42**(8), 147 (2018)
- F. Loukil, C. Ghedira-Guegan, K. Boukadi, A.N. Benharkat, in *International Conference on Web Information Systems Engineering* (Springer, 2018), pp. 68–78

11. X. Chen, X. Wang, K. Yang, in *2019 IEEE International Conference on Big Data (Big Data)* (2019), pp. 5469–5473
12. A. Alnemari, S. Arodi, V.R. Sosa, S. Pandey, C. Romanowski, R. Raj, S. Mishra, in *International Conference on Critical Infrastructure Protection* (Springer, 2018), pp. 113–125
13. G. Zyskind, O. Nathan, A.. Pentland, in *2015 IEEE Security and Privacy Workshops* (2015), pp. 180–184
14. C. Clifton, T. Tassa, in *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)* (IEEE, 2013), pp. 88–93
15. F. Kohlmayer, F. Prasser, C. Eckert, A. Kemper, K.A. Kuhn, in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing* (IEEE, 2012), pp. 708–717
16. F. Prasser, F. Kohlmayer, R. Lautenschlaeger, K.A. Kuhn, in *AMIA Annual Symposium Proceedings*, vol. 2014 (American Medical Informatics Association, 2014), vol. 2014, p. 984
17. K. El Emam, F.K. Dankar, R. Issa, E. Jonker, D. Amyot, E. Cogo, J.P. Corriveau, M. Walker, S. Chowdhury, R. Vaillancourt, et al., *Journal of the American Medical Informatics Association* **16**(5), 670 (2009)
18. N. Li, W.H. Qardaji, D. Su, CoRR, abs/1101.2604 **49**, 55 (2011)
19. M.J.M. Chowdhury, A. Colman, J. Han, M.A. Kabir, in *Proceedings of the 51rd Hawaii International Conference on System Sciences* (2018), pp. 1–10. DOI 10.24251/HICSS.2018.594. URL <http://hdl.handle.net/10125/50483>
20. M.J.M. Chowdhury, A. Colman, M.A. Kabir, J. Han, P. Sarda, in *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (IEEE, 2019), pp. 327–333
21. M.J.M. Chowdhury, A. Colman, J. Han, M.A. Kabir, in *Proceedings of the Australasian Computer Science Week Multiconference* (2018), pp. 1–10
22. C. Dwork, in *International conference on theory and applications of models of computation* (Springer, 2008), pp. 1–19
23. J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, S. Martínez, *The VLDB Journal* **23**(5), 771 (2014)
24. A. Hutchinson. Facebook outlines new differential privacy framework to protect user information in shared datasets. <https://www.socialmediatoday.com/news/facebook-outlines-new-differential-privacy-framework-to-protect-user-inform/579167/> (June 3, 2020). [Online; accessed 13-09-2020]
25. Ú. Erlingsson, V. Pihur, A. Korolova, in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (2014), pp. 1054–1067
26. Uber becomes the latest company to embrace differential privacy. <https://iapp.org/news/a/uber-becomes-the-latest-company-to-embrace-differential-privacy/> (Jul 14, 2017). [Online; accessed September 13, 2020]
27. T. Zhu, G. Li, W. Zhou, S.Y. Philip, *Differential privacy and applications*, vol. 69 (Springer, 2017)
28. S.P. Kasiviswanathan, H.K. Lee, K. Nissim, S. Raskhodnikova, A. Smith, *SIAM Journal on Computing* **40**(3), 793 (2011)
29. J.C. Duchi, M.I. Jordan, M.J. Wainwright, in *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (IEEE, 2013), pp. 429–438
30. C. Dwork, A. Roth, et al., *Foundations and Trends® in Theoretical Computer Science* **9**(3–4), 211 (2014)
31. T. Poongodi, R. Krishnamurthi, R. Indrakumari, P. Suresh, B. Balusamy, in *A Handbook of Internet of Things in Biomedical and Cyber Physical System* (Springer, 2020), pp. 245–273
32. L. Cooper. Medical-grade devices vs. consumer wearables. <https://www.electronicsspecifier.com/products/wearables/medical-grade-devices-vs-consumer-wearables> (2nd July 2019). [Online; accessed September 16, 2020]
33. K. Mammadzada, M. Iqbal, F. Milani, L. García-Bañuelos, R. Matulevičius, in *International Conference on Business Process Management* (Springer, 2020), pp. 19–34
34. C. Okoli, *Communications of the Association for Information Systems* **37**(1), 43 (2015)
35. A. Fink, *Conducting research literature reviews: From the internet to paper* (Sage publications, 2019)
36. D. Moher, D.G. Altman, A. Liberati, J. Tetzlaff, *Epidemiology* **22**(1), 128 (2011)
37. C. Lin, Z. Song, H. Song, Y. Zhou, Y. Wang, G. Wu, *Journal of medical systems* **40**(4), 97 (2016)
38. C. Lin, P. Wang, H. Song, Y. Zhou, Q. Liu, G. Wu, *Annals of Telecommunications* **71**(9-10), 465 (2016)
39. M. Hadian, X. Liang, T. Altuwaiyan, M.M. Mahmoud, in *2016 IEEE Global Communications Conference (GLOBECOM)* (IEEE, 2016), pp. 1–6
40. K. Prema, A. Sriharsha, *Technology* **8**(3), 11 (2017)
41. J. Zhang, X. Liang, Z. Zhang, S. He, Z. Shi, in *GLOBECOM 2017-2017 IEEE Global Communications Conference* (IEEE, 2017), pp. 1–6
42. S. Han, S. Zhao, Q. Li, C.H. Ju, W. Zhou, *IEEE Transactions on Information Forensics and Security* **11**(9), 1940 (2015)

43. Z. Guan, Z. Lv, X. Du, L. Wu, M. Guizani, Future Generation Computer Systems **98**, 60 (2019)
44. J.W. Kim, J.H. Lim, S.M. Moon, H. Yoo, B. Jang, in *2019 IEEE International Conference on Consumer Electronics (ICCE)* (IEEE, 2019), pp. 1–4
45. J.W. Kim, J.H. Lim, S.M. Moon, B. Jang, IEEE Transactions on Consumer Electronics **65**(3), 369 (2019)
46. J.H. Lim, J.W. Kim, Journal of the Korea Society of Computer and Information **24**(9), 43 (2019)
47. H. Ren, H. Li, X. Liang, S. He, Y. Dai, L. Zhao, Sensors **16**(9), 1463 (2016)
48. J.W. Kim, B. Jang, H. Yoo, PloS one **13**(11) (2018)
49. Z. Zhang, B. Han, H.C. Chao, F. Sun, L. Uden, D. Tang, IEEE Access **7**, 104045 (2019)
50. J. Steil, I. Hagedeit, M.X. Huang, A. Bulling, in *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (2019), pp. 1–9
51. Y. Zhang, Y. Qu, L. Gao, T.H. Luan, X. Zheng, S. Chen, Y. Xiang, IEEE Access **7**, 166593 (2019)
52. N. Saleheen, S. Chakraborty, N. Ali, M.M. Rahman, S.M. Hossain, R. Bari, E. Buder, M. Srivastava, S. Kumar, in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2016), pp. 706–717
53. E. Bozkir, O. Günlü, W. Fuhl, R.F. Schaefer, E. Kasneci, arXiv preprint arXiv:2002.08972 (2020)
54. A. Ukil, A.J. Jara, L. Marin, Sensors **19**(12), 2733 (2019)
55. L. Fan, L. Xiong, in *Proceedings of the 21st ACM international conference on Information and knowledge management* (2012), pp. 2169–2173
56. J. Wang, R. Zhu, S. Liu, IEEE Access **6**, 6487 (2018)
57. V. Rastogi, S. Nath, in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data* (2010), pp. 735–746
58. E. Shi, T.H. Chan, E. Rieffel, R. Chow, D. Song, in *Proc. NDSS*, vol. 2 (Citeseer, 2011), vol. 2, pp. 1–17
59. Y. Cao, L. Xiong, M. Yoshikawa, Y. Xiao, S. Zhang, Proceedings of the VLDB Endowment **11**(12), 2090 (2018)
60. Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, K. Ren, in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications* (IEEE, 2016), pp. 1–9
61. R. Gao, X. Ma, in *2018 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications (ISPA/IUCC/BDCloud/SocialCom/SustainCom)* (IEEE, 2018), pp. 737–743
62. L. Fan, L. Xiong, IEEE Transactions on knowledge and data engineering **26**(9), 2094 (2013)
63. G. Kellaris, S. Papadopoulos, X. Xiao, D. Papadias, Proc. VLDB Endow. **7**(12), 1155–1166 (2014)
64. T.T. Nguyêñ, X. Xiao, Y. Yang, S.C. Hui, H. Shin, J. Shin, arXiv preprint arXiv:1606.05053 (2016)
65. M. Yang, Improving privacy preserving in modern applications. Tech. rep., Deakin University (2019)
66. T. Asikis, E. Pournaras, Future Generation Computer Systems **109**, 488 (2020)
67. C. Luo, X. Liu, W. Xue, Y. Shen, J. Li, W. Hu, A.X. Liu, IEEE/ACM Transactions on Networking **27**(1), 361 (2019)
68. Wikipedia contributors. Haar wavelet — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Haar\\_wavelet&oldid=950816833](https://en.wikipedia.org/w/index.php?title=Haar_wavelet&oldid=950816833) (2020). [Online; accessed August 19, 2020]
69. Wikipedia contributors. Kalman filter — Wikipedia, the free encyclopedia (2020). URL [https://en.wikipedia.org/w/index.php?title=Kalman\\_filter&oldid=974917947](https://en.wikipedia.org/w/index.php?title=Kalman_filter&oldid=974917947). [Online; accessed September 3, 2020]
70. Wikipedia contributors. Kullback–leibler divergence — Wikipedia, the free encyclopedia. [https://en.wikipedia.org/w/index.php?title=Kullback%25E2%2580%2593Leibler\\_divergence&oldid=976499574](https://en.wikipedia.org/w/index.php?title=Kullback%25E2%2580%2593Leibler_divergence&oldid=976499574) (2020). [Online; accessed 23-September-2020]
71. A. Reiss, D. Stricker, in *2012 16th International Symposium on Wearable Computers* (2012), pp. 108–109
72. A. Fleury, M. Vacher, N. Noury, IEEE Transactions on Information Technology in Biomedicine **14**(2), 274 (2010)
73. A. Fleury, N. Noury, M. Vacher, International Journal of E-Health and Medical Communications (IJEHMC) **2**(1), 17 (2011)
74. M. Kassner, W. Patera, A. Bulling, in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Association for Computing Machinery, New York, NY, USA, 2014), UbiComp ’14 Adjunct, p. 1151–1160. DOI 10.1145/2638728.2641695. URL <https://doi.org/10.1145/2638728.2641695>
75. I.C. Yeh, K.J. Yang, T.M. Ting, Expert Systems with Applications **36**(3), 5866 (2009)
76. Influenza(flu) dataset. <https://www.cdc.gov/flu/> (May 15, 2020). [Online; accessed May 17, 2020]
77. Traffic dataset. <http://www.its.washington.edu/>. [Not Available; accessed May 17, 2020]
78. Unemployment dataset. <https://research.stlouisfed.org/>. [Online; accessed May 17, 2020]
79. Traffic dataset. <http://www.transguide.dot.state.tx.us/>. [Not available;

- accessed May 17, 2020]
80. Taxi trajectory prediction. <https://www.kaggle.com/c/pkdd-15-predict-taxi-service-trajectory-i/data> (2015). [Online; accessed May 17, 2020]
  81. World cup dataset. <https://ita.ee.lbl.gov/html/contrib/WorldCup.html>. [Not available; accessed 17-05-2020]
  82. T. Brinkhoff, *GeoInformatica* **6**(2), 153 (2002)
  83. differential-privacy. [https://github.com/topics/differential-privacy?fbclid=IwAR0fvaB4kSAr4-C7f7fVMevVvy9-mykJcWPpb4-kbRmA\\_hlqpnFDfsy0UdY](https://github.com/topics/differential-privacy?fbclid=IwAR0fvaB4kSAr4-C7f7fVMevVvy9-mykJcWPpb4-kbRmA_hlqpnFDfsy0UdY). [Online; accessed October 29, 2020]
  84. V. Davis. Google open sources their differential privacy library to help protect user's private data. <https://hub.packtpub.com/google-open-sources-their-differential-privacy-library-to-help-protect-users-private-data/> (September 6, 2019). [Online; accessed October 29, 2020]
  85. H. Li, Y. Dai, X. Lin, in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)* (IEEE, 2015), pp. 602–608
  86. J. Zhao, Y. Chen, W. Zhang, *IEEE Access* **7**, 48901 (2019)
  87. F.K. Dankar, K. El Emam, in *Proceedings of the 2012 Joint EDBT/ICDT Workshops* (Association for Computing Machinery, New York, NY, USA, 2012), EDBT-ICDT '12, p. 158–166. DOI 10.1145/2320765.2320816. URL <https://doi.org/10.1145/2320765.2320816>
  88. T. Zhu, P. Xiong, G. Li, W. Zhou, *IEEE Transactions on Information Forensics and Security* **10**(2), 229 (2014)
  89. B. Yang, I. Sato, H. Nakagawa, in *Proceedings of the 2015 ACM SIGMOD international conference on Management of Data* (2015), pp. 747–762
  90. X. Xiao, G. Wang, J. Gehrke, *IEEE Transactions on Knowledge and Data Engineering* **23**(8), 1200 (2011)
  91. H. Wang, Z. Xu, *Knowledge-Based Systems* **122**, 167 (2017)
  92. X. He, G. Cormode, A. Machanavajjhala, C.M. Procopiuc, D. Srivastava, *Proceedings of the VLDB Endowment* **8**(11), 1154 (2015)
  93. R. Sarathy, K. Muralidhar, *Trans. Data Privacy* **4**(1), 1 (2011)
  94. M.J.M. Chowdhury, M.S. Ferdous, K. Biswas, N. Chowdhury, A. Kayes, M. Alazab, P. Watters, *IEEE Access* **7**(1), 167930 (2019)
  95. Z. Xiong, Y. Zhang, D. Niyato, P. Wang, Z. Han, *IEEE Communications Magazine* **56**(8), 33 (2018)
  96. M. Banerjee, J. Lee, K.K.R. Choo, *Digital Communications and Networks* **4**(3), 149 (2018)
  97. A. Alnemari, S. Arodi, V.R. Sosa, S. Pandey, C. Romanowski, R. Raj, S. Mishra, in *International Conference on Critical Infrastructure Protection* (Springer, 2018), pp. 113–125
  98. M.J.M. Chowdhury, M.S. Ferdous, K. Biswas, N. Chowdhury, V. Muthukumarasamy, *Knowledge Engineering Review* **35**, 22 (2020)
  99. M. Mettler, in *2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom)* (IEEE, 2016), pp. 1–3
  100. M.U. Hassan, M.H. Rehmani, J. Chen, *Journal of Parallel and Distributed Computing* **145**, 50 (2020)
  101. Y. Zhao, J. Zhao, J. Kang, Z. Zhang, D. Niyato, S. Shi, K.Y. Lam, *IEEE Internet of Things Journal* **8**(11), 8865 (2021)
  102. L.M. Han, Y. Zhao, J. Zhao, arXiv preprint arXiv:2006.04693 (2020)
  103. T. Zhu, S.Y. Philip, in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)* (IEEE, 2019), pp. 1601–1609
  104. K.M.P. Shrivastva, M. Rizvi, S. Singh, in *2014 International Conference on Computational Intelligence and Communication Networks* (IEEE, 2014), pp. 776–781

## Appendix A: Appendix: Paper Search and Review

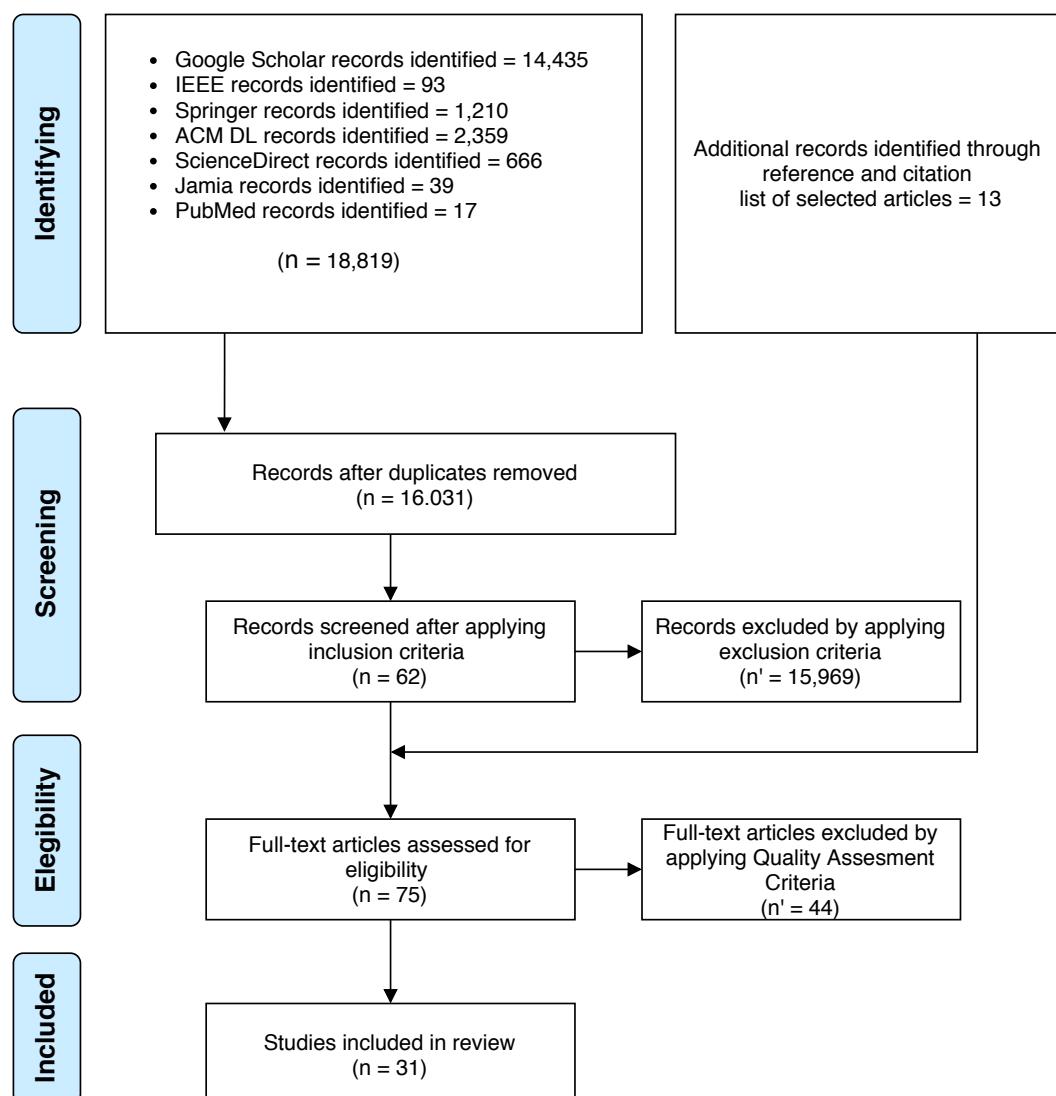


Fig. 18: PRISMA flow diagram

Table 13: Retrieved papers using logical AND &amp; OR

<b>Keywords (using logical AND &amp; OR)</b>	<b>Number of retrieved papers</b>						
	<b>Google Scholar</b>	<b>IEEE</b>	<b>ACM DL</b>	<b>Science Direct</b>	<b>Springer</b>	<b>JAMIA</b>	<b>PubMed</b>
(Wearable OR "privacy preserving") AND ("data publishing" AND "differential privacy")	3170	45	157	132	288	0	4
("Temporal data" OR "wearable data") AND ("publish using" OR "publish under") AND ("differential privacy")	0	0	0	27	2	0	0
(Wearable OR Medical OR Health) AND "data privacy" AND (using OR under) AND "differential privacy"	5920	0	311	188	472	14	3
("Wearable" OR "Wearable devices generated") AND "data privacy" AND (using OR under) AND "differential privacy"	901	0	63	54	66	0	0
((Review OR Survey OR SLR OR "Literature Review") on AND ("Wearable Data" OR "Wearable devices data")) AND ((("Publishing using" OR "data publishing") AND ("differential privacy")	4	0	177	1	0	0	0
(Wearable OR "privacy preserving") AND (("data publishing" OR "publishing using") AND "differential privacy")	3170	45	157	209	289	0	10
(Wearable data) AND (publish under OR publishing using) AND "differential privacy"	1270	3	1494	55	93	25	0
<b>Total retrieved:</b>	<b>14435</b>	<b>93</b>	<b>2359</b>	<b>666</b>	<b>1210</b>	<b>39</b>	<b>17</b>