Wireless Personal Communications (2022) 126:2379–2401
https://doi.org/10.1007/s11277-021-09323-0

A Deep Learning-Based Privacy-Preserving Model for Smart Healthcare in Internet of Medical Things Using Fog Computing

Syed Atif Moqurrab[1] · Noshina Tariq[2] · Adeel Anjum[1,3] · Alia Asheralieva[3] · Saif U. R. Malik[4] · Hassan Malik[5] · Haris Pervaiz[6] · Sukhpal Singh Gill[7]

Accepted: 4 November 2021 / Published online: 30 August 2022

Abstract

With the emergence of COVID-19, smart healthcare, the Internet of Medical Things, and big data-driven medical applications have become even more important. The biomedical data produced is highly confidential and private. Unfortunately, conventional health systems cannot support such a colossal amount of biomedical data. Hence, data is typically stored and shared through the cloud. The shared data is then used for different purposes, such as research and discovery of unprecedented facts. Typically, biomedical data appear in textual form (e.g., test reports, prescriptions, and diagnosis). Unfortunately, such data is prone to several security threats and attacks, for example, privacy and confidentiality breach. Although significant progress has been made on securing biomedical data, most existing approaches yield long delays and cannot accommodate real-time responses. This paper proposes a novel fog-enabled privacy-preserving model called δ_r *sanitizer*, which uses deep learning to improve the healthcare system. The proposed model is based on a Convolutional Neural Network with Bidirectional-LSTM and effectively performs Medical Entity Recognition. The experimental results show that δ_r sanitizer outperforms the state-of-the-art models with 91.14% recall, 92.63% in precision, and 92% F1-score. The sanitization model shows 28.77% improved utility preservation as compared to the state-of-the-art.

Keywords Internet of Things · Fog computing · Machine learning · Smart healthcare · Privacy · Sanitization

1 Introduction

Digital technologies, such as the Internet of Medical Things (IoMT), big-data analytics, 5G, and Artificial Intelligence (AI), have revolutionized critical diseases and medical illness prevention, monitoring, and treatment [1]. IoMT with 5G-connected devices supplies medical aids with new and advanced facilities. Patients and medical staff (doctors,

✉ Sukhpal Singh Gill
s.s.gill@qmul.ac.uk

Extended author information available on the last page of the article

nurses, and health organizations) can use their mobile devices to remain in contact, lowering the rate of physically hospitalizing a patient. As such, the IoMT-based platforms will help gather health-related data and provide access to health organizations to prevent, control, and mitigate the spread of viral infections, e.g., COVID-19. In an IoMT smart healthcare ecosystem, such massive data is synchronized to the cloud for storage and analysis [2]. The cloud-based smart healthcare ecosystem comprises of sensor layer and the cloud layer. The sensor layer provides the patients' credentials and sensory data, which is then stored in the cloud [3]. This data may be shared with health organizations, families, and authorized parties for research purposes. It also enables the healthcare facilities to be delivered to the isolated areas on time, at reasonable prices. Smart healthcare systems use different devices, such as wireless sensors, cameras, and controllers, to allow patients' automated recognition, awareness of the right medication, and serious initial signals to detect health decline (seizure, heart failure, test results, e.g., COVID-19 test and temperature measurements).

As anticipated in [4], current hospital-home healthcare systems will turn into only home-centred systems by the year 2030. For example, the current COVID-related emergency has already confined people/patients to their homes. As a result, present off-line healthcare systems are re-shaping accordingly into digital smart healthcare systems [5]. To meet these evolutionary changes, advanced healthcare infrastructures and technologies must be taken into account. However, such technological shift en route to pervasive smart healthcare systems brings upon new challenges, such as security, efficiency in terms of latency and energy consumption, inter-operability, mobility, reliability and privacy [6].

In the current IoT cloud infrastructure, the mobility of things needs further improvement. It covers only hospital/building premises, which causes poor scalability and efficiency [7]. Moreover, connected devices and massive data's pervasiveness bring congestion in the network and unwanted delay in cloud-based smart healthcare infrastructure. Thus, the cloud computing paradigm is not suitable for such infrastructures, leading to fatal consequences. Therefore, fog computing can be introduced between sensor devices and the cloud layer to meet delay-sensitive health application needs. It provides cloud-like services at the edge of the network. It works as an intermediary computation layer, which provides scalability, low latency, low power consumption, seamless mobility, and many other advantages, e.g., as summarized in [8]. However, both cloud and fog introduce new security threats to the smart-heath domain. These attacks include (but are not limited to) confidentiality, integrity, anonymity, privacy, and data freshness.

1.1 Motivation

The electronic form of medical data (AKA Electronic Medical Record (EMR) or Electronic Health Record (EHR)) is growing massively. It is generally classified into structured (i.e., statistical databases, tables) and unstructured data (i.e., text, videos, images, and voice). For the safety of structured medical data different security measures are used, for example *k*-anonymity, *l*-diversity, *t*-closeness, Differential Privacy (DP), and relaxed form of (DP) [9]. Since much clinical information is text-based, it is inevitable to discover the solutions to protect such data. Redaction is the initial step to mask or remove any secret information from the piece to preserve medical data. It completely changes the meaning and degrades the use of information. Thus, there is another way called "sanitization," which converts the most important phrases to the least important ones, such as "Corona", can be masked by "Virus." The idea is extensively utilized to achieve safety measures for unstructured data. Many recognized methods used for word-based articles are according to the analytical concept of data. However, many other restrictions should be considered. To detect the clinical equipment automatically from the collection, this concept needs simple procedures. Consequently, several Machine Learning (ML) techniques have been used to detect medical entities, such as diagnosis, testing, and treatments.

The Deep Learning (DP) models are the most dominant ones for unstructured data classification [10, 11]. However, more intelligent designs must be discovered to improve efficiency. In most of the existing ML and DP-based solutions, privacy breaches along with utility issues are inevitable. Therefore, it is important to design such a model, which may handle both issues effectively. Then, redaction-based techniques were used to overcome issues related to manual anonymization [12]. The redaction of unstructured data is a process of removal of sensitive terms from raw medical data being sensed or acquired. However, removing sensitive terms may also change the semantics of the underlying document. It may also affect the quality of the document resulting in less usability and lead to a privacy breach.

1.2 Our Contributions

The Internet of Things (IoT) is a rapidly developing technology that seeks to provide ubiquitous access (at any time and from any location) to a wide variety of devices through the Internet. It serves as the foundation for various smart applications, including automation and monitoring in smart healthcare systems. Various technologies, such as fog computing, contributes significantly to the concept of vast and intelligent connectivity. It helps boost quality and dependability by offering innovative computing options and resource planning [13, 14]. Cloud services may help the IoT get inexpensive on-demand solutions for large data storage and heavy processing. Unfortunately, there are still unsolved problems in cloud-based IoT applications, such as high capacity client access, fluctuating delay, safety, and less mobility and location awareness [15]. Applications such as real-time health monitoring, in particular, is highly delay-sensitive to cloud facilities. Fog computing, which provides various services and numerous resources to end-users at the edge of the network, has been developed to solve these issues. It relies on local networks rather than a central cloud architecture to create specialised channels. It improves the end-user Quality of Service (QoS) and user experience, guaranteeing decreased service latency. In this paper, we proposed a fog computing-based privacy model for smart healthcare using deep learning called δ_r ***sanitizer***, which considers a smart healthcare system realized in the fog network to reduce latency for delay-sensitive medical applications. The proposed scheme can handle heterogeneous data collected from heterogeneous networks having different data structures and types, such as numeric, alphanumeric, and textual. The proposed MER minimizes latency and energy consumption at the sensor (node) level. The suggested model has a wide range of applications. Clinical entity detection offers a wide variety of applications in the biomedical sector, including proteins, genes detection, illnesses, and medication chemical formulae. It aids in optimising search queries, the interpretation of clinical reports, and the protection of biomedical data's privacy [16]. Biomedical data privacy is a relatively new field of study. It enables institutions, such as hospitals, to share medical data (in a secure manner) with the research groups without jeopardising patient confidentiality [17]. While sharing medical data with researchers may help improve healthcare and provide better treatments for illnesses, it cannot be released in its entirety to preserve individuals' privacy. To maintain confidentiality, it is necessary to identify clinical entities accurately. Confidentiality is preserved when clinical entities (disease, test, and therapy) are accurately identified and sanitised (generalised). The main contributions of the paper are as follows:

(1) We present a novel fog-enabled privacy-preserving framework called Deep Privacy to mitigate latency and energy-consumption issues at the node-level by using deep-learning-based Medical Entity Recognition (MER) scheme with enhanced recognition accuracy by combining local and global contextual representation.
(2) The proposed framework improves the privacy of unstructured biomedical data by enforcing medical entity sanitization, called δ_r-sanitization, a variant of sanitization mechanisms that not only sanitizes the data but also preserves its utility at maximum.

The rest of the paper is organized as follows: Sect. 2 discusses the related work. In Sect. 3, we detail the proposed framework. Sect. 4 discusses medical entity recognition and Sect. 5introduces enhanced sanitization model. In Sect. 6, the numerical analysis and results are discussed. In Sect. 7, the experimental analysis and discussion is presented. And lastly, the conclusion and future work is given in Sect. 8.

2 Related Work

Recently, contextual embedding-based models have been proposed using both character-level and token-level representation to improve MER accuracy, for instance Word2Vec [18], GloVec [19], ELMO [20], Bert [21], and Bio-Bert [22]. However, These models require rigorous computational resources and cause a high processing cost. Therefore, there is still a need to explore the architecture of deep learning-based models for achieving high recognition accuracy. Initially, Information theoretic-based models were proposed for document sanitization. One of these models is the local Information Content (IC) that is used for sanitization. However, this model is biased towards the local high-frequency occurrence of the terms [23].

Sanchez et al. proposed a novel privacy-preserving model, based on sanitization, to overcome issues related to IC models in [24]. However, it has non-monotonically behavior for available medical concepts‘ taxonomy. It also shows language ambiguity [25]. Saha et al. [26] discussed the use of fog layers in IoT-based healthcare systems and their usage in dealing with EMRs. Zhao et al. [27] propounded a privacy-preserving data aggregation scheme for edge-based VANETs. It reduced computing and communication overhead at the node level and also released the communication pressure on the edge. However, there is still room for reducing computation overhead on the cloud center. Similarly, Dong et al. [28] investigated an edge-based healthcare system in IoMTs to minimize the system-wide cost in the edge-based healthcare systems. However, MUs lead to deficient wireless channels and computation resources.

For VANETs, Sui et al. [29] presented an edge-based privacy-preserving data downloading technique. However, cooperative downloading and a lightweight and cryptography-based incentive are required for resource-constrained devices. Wang et al. [30] proposed a scheme using fog-based content transmission and collective filtering for vehicles. Bouchelaghem and Omer [31] proposed a privacy-preserving mechanism using a pseudonym changing strategy for VNs, where they can communicate autonomously. However, the simulation area used in this research is small and not suitable for big cities. Guan et al. [32] proposed a public auditing scheme for fog-to-cloud data storage integrity and privacy. However, the scheme requires extra devices to mitigate computation and throughput overheads. Moreover, it may lead to unwanted homomorphic encryption.

2.1 Summary of Related Work

To summarize, ML-based mechanisms require immense pre-processing and parameter tuning, which can be improved using deep learning-based models. However, it is still needed to explore various architectures, such as CNN, LSTM. Not much work has been done in this area, so far, to the best of our knowledge. Most of the proposed mechanisms either use local or global context for MER. Having said that, if both the local and the global contexts are conjoined, the detection accuracy may be improved (as in our case). Despite that, most of the mechanisms focus only on word embedding rather than exploring promising (above mentioned) architectures to improve accuracy. That is why significant improvement has not been observed. Also, both the recognition and the sanitization mechanisms are not implemented and used as a single technique. Therefore, it is inevitable to propose a single complete architecture that considers both. To this end, we already proposed N-sanitization [17] without using the concept of deep learning. We analyzed our previous technique also and came up with a fog computing-based privacy model using deep learning. Table 1 compares the proposed model with existing works.

3 System Model and Architecture

This section proposes a novel framework called Deep Privacy that enables the automatic detection and sanitization of the medical entities for medical records (e.g., sensors‘ data, reports, and medicines prescriptions). Since these documents hold the patients’ data, it becomes mandatory to sanitize before printing and transferring. Therefore, sanitization-based approaches were proposed [17]. In sanitization, the sensitive and most-concerned

Table 1 Notation description

Symbols	Meanings
σ	Sigmoid function
$\otimes$	Element-wise product
$Tanh$	Tangent-hyperbolic function
δ	Privacy threshold
δ_r	Random privacy threshold
$SenT_i$	Single sensitive term
$SenT$	Sanitized terms
C_T	Clinical taxonomy
C_T_n	Lenght of clinical taxonomy
C_T_{gen}	Most generalized clinical term
$sanitize(SenT)$	Generalized/sanitized sensitive term
MKB	Clinical terms taxonomy

Table 2 Comparison of δ_r sanitizer with related works

Reference	Cloud/Fog	Deep learning	IoT	Detection	Sanitization	Overhead			
						Energy	Latency	Computation	Storage
[18]	×	✓	×	✓	×	✓	✓	✓	✓
[22]	×	✓	×	✓	×	✓	✓	✓	✓
[27]	Fog	×	✓	×	✓	×	×	×	×
[28]	Fog	×	✓	×	×	×	×	×	×
[31]	Cloud	×	✓	×	✓	×	✓	×	×
[22]	×	×	×	×	✓	✓	×	×	✓
[25]	×	×	×	✓	✓	✓	✓	✓	✓
Proposed	Fog	✓	✓	✓	✓	×	×	×	×

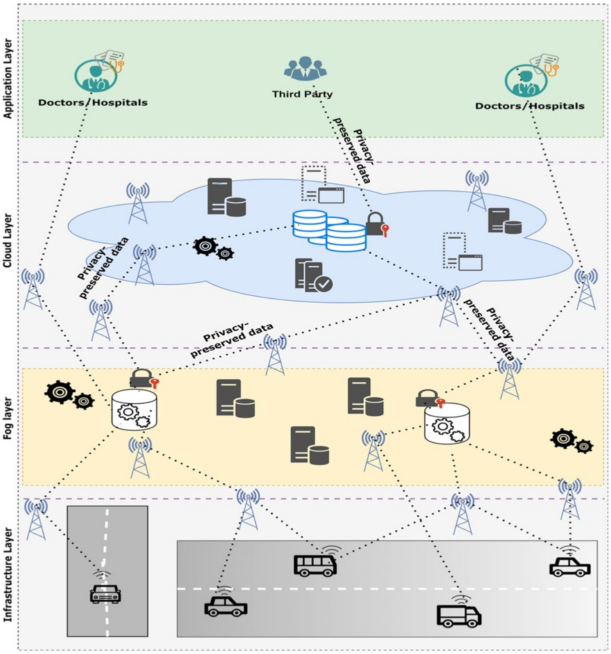


Fig. 1 Proposed architecture of fog-enabled δ_r-Sanitizer

terms are masked with generic and less certain terms. In this way, the unstructured medical data's privacy may be preserved against privacy breaches and threats (Table 2). Figure 1 shows the proposed architecture of fog-enabled Deep Privacy framework is divided into four layers. In addition, Fig. 2 shows the activity diagram (overall flow) of the proposed framework. The detailed activity of the proposed model (and layers) is discussed below:

(1) *Application layer* This layer targets the users of the medical data that has been sensed from the sensor devices. The users may be doctors, paramedical staff, the patient himself, and his family. Third parties, such as different organizations conducting research, are also a part of this layer. However, they access sanitized data from the cloud, unlike the rest of the users, accessing the fog layer's data.

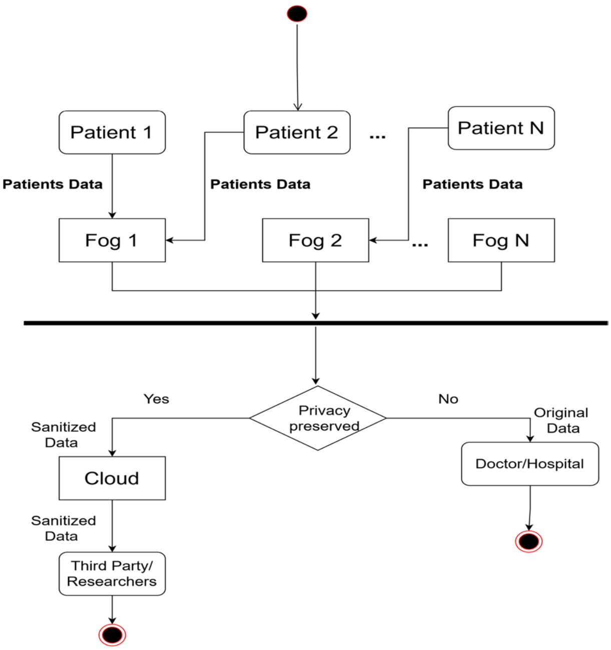


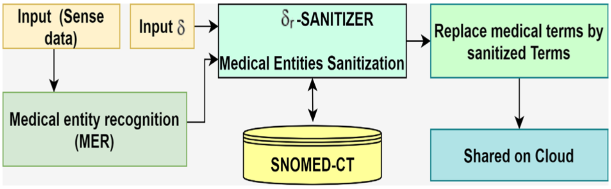
Fig. 2 Activity diagram of proposed framework

(2) *Cloud layer* The cloud layer receives privacy-preserved (sanitized) data from the fog layer and provides permanent storage for that. Third parties, such as researchers, publishers, and pharmaceutical organizations, may request the cloud layer data.

(3) *Fog layer* The fog layer is introduced to minimize latency for delay-critical real-time applications (e.g., in the case of health emergencies), reduce power consumption, and avoid congestion in the network backhaul [33]. The fog layer acts as a control layer, responsible for medical data recognition and privacy preservation. It comprises two main modules: i) MER module to recognize unstructured data and ii) δ_r-sanitized to sanitize the data, as shown in Fig. 3. It depicts that detected sensitive terms are further sanitized in the fog node and forwarded to the cloud for permanent storage. This sanitized data is accessible by third parties (researchers) via the cloud for research purposes.

The data received from the infrastructure layer are summarized in reports. The medical practitioner generated prescriptions are saved transiently on this layer. In this way, before sending the entire data to the cloud, only summarised data is forwarded to save bandwidth and minimize latency. The medical correspondents and patients can access that data directly for real-time responses from the fog. The data is sanitized before sending it to the cloud layer for permanent storage. After this, the sanitized data may be shared with third

Fig. 3 Flow of sequence in δ_r-Sanitizer


Input (Sense data)
Input δ
δr-SANITIZER
Medical Entities Sanitization
Replace medical terms by sanitized Terms
Medical entity recognition (MER)
SNOMED-CT
Shared on Cloud

parties for research purposes. This data is secured using the Deep Privacy mechanism. Suppose the data security is breached during the transition from the fog layer to the cloud. In that case, the adversaries cannot get the real unsanitized data. However, the security of data sent from the infrastructure layer to the fog is beyond this paper's scope. We assumed that the data sensed from a device and transit to the fog layer was secured.

Firstly, the unstructured data (sensor and textual data) is forwarded to the MER module for recognizing medical entities. To recognize them accurately, a new concept is considered, which is the combination of local context using CNN and global context using LSTM with CRF. Secondly, the recognized entities are shipped to δ_r-Sanitizer module for sanitization along with δ. δ is a threshold that is used to control the privacy and utility trade-off. For instance, while preserving privacy in the sanitization process, a term is generalized. For example, the term "COVID" may be generalized as "Virus." Thus, there is a trade-off between privacy and utility. The more is the generalization. The less is the utility [9]. Thereby, δ is used to balance between them. If the value of δ is high, the privacy will increase, and the utility will decrease, and vice versa. Therefore, we have not set a value of δ. We left it to the data-holder to set the value according to their requirements. After receiving inputs from the MER module and δ from the requester, δ_r-Sanitizer sanitizes the recognized entity using SNOMED-CT (knowledge-base [34]) for generalization. Lastly, the medical terms are replaced by sanitized terms and shared on the cloud.

(4) *Infrastructure layer* This layer consists of smart sensor devices that are attached to a patient on the move and at rest. These sensors keep observing patients‘ vitals and keep on sending the data to the fog layer. The data is sent to the healthcare correspondents. In case of any abnormal and emergency, the response is sent back to the patient or their family or caretakers.

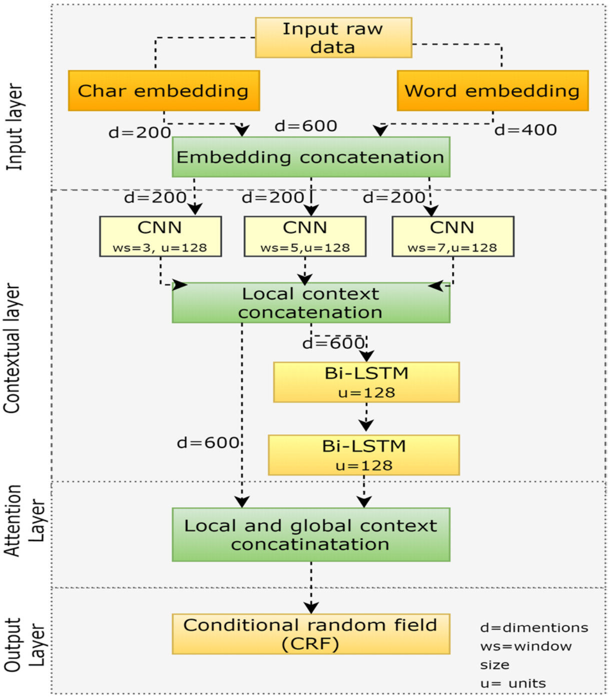
4 Medical Entity Recognition

For medical entity recognition, we deployed a variant of RNN, called Bi-LSTM, an alternative to RNN with CRF. It is suitable for such issues in terms of efficiency, as mentioned in [35]. Moreover, it can avoid gradient varnishing and hold the long-distance dependencies of the related information. It emphasizes more on the information, which is related to the global context than the local ones. Hence, a new design is considered using CNN with Bi-LSTM to grab local and global context more efficiently. Figure 4 illustrates the MER model with four layers. The components of the model are given below:

Input layer The eventual occurrence of each word is recorded using two separate token presentations called; token-level and character-level representations. This representation is more often called word embedding and character embedding. The token saves the relative data in the vector space, and the grammatical information is recorded by character-level representation. The token-level representation uses the already utilized Consecutive Bags of Words (CBoW) and skip-gram models [35]. The character-level representation uses bidirectional LSTM. Consequently, the token- and the character-level representations are combined to couple every word. This combined representation is transferred to the contextual layer.

Contextual layer This layer produces local context for every token representation utilizing CNN. The token representation from the input layer is taken by CNN and generates contextual information. Three different window sizes (i.e., 3, 5, 7) are used to achieve an efficient local context. These window sizes generate different contextual information,

Fig. 4 The proposed deep neural networks-based MER model



which plays an important role in efficiency by transiting from confined to a wider context [10]. The local context of the said windows is coupled together for efficacy. After this, the particular representation is transferred to the bi-LSTM layer to acquire the global context. Based on the integrated CNN data, the bi-LSTM generates global data in sequence. For instance, a phrase $P = (p_1, p_2, p_3, \dots, p_n)$ with every text CNN represents local data as $M = (m_1, m_2, m_3, \dots, m_n)$. This relative data is taken by bi-LSTM as an input and produces global presentation as $GL = (gl_1, gl_2, gl_3, \dots, gl_n)$, where $GL^n = [GL^N_{fn}, GL^N_{bn}]^N$ is a fusion of output for all the backward and forward passes of LSTM.

Attention layer This layer couples CNN-generated local context representation with the global context representation generated by Bi-LSTM. After the coupling of the context, it forwards it to the output layer.

Output layer The output layer takes the sequence $GL = (gl_1, gl_2, gl_3, \dots, gl_n)$ from attention layer as an input and uses CRF to predict the best possible label sequence for each token $S = (s_1, s_2, s_3, \dots, s_n)$. For example, an input training data-set GL and all the variables of the CRF model (θ) can be computed by increasing the log-likelihood by using Eq. 1:

$$C(\theta) = \sum_{(p,s)\in GL} log(c_p(s|gl, \theta)) \tag{1}$$

where s denotes order label for string of tokens denoted by P, c_p, is a conditional occurrence of s, given, that, P and θ. If we assume that $P_\theta(gl, s)$ is the result of the label sequence s for every token in the phrase, then, the standardization of $P_\theta(gl, s)$ is an approximation of conditional probability c_p. To get the highest spot of labels, which have the closest mate, the CRF design combines emission matrix M along with a transition matrix N for the calculation of the result of the label sequence $P_{\theta}(gl, s)$ as follows (Eq. 2):

$$P_{\theta}(gl, s) = \sum_{q=1}^{n} (M_{sq,q+Q_{sq-1,sq}}) \tag{2}$$

In $M_{sq,q}$, q represents the probability of token (word) gl_q with the label s_q. The $Q_{sq-1,sq}$ represents probability of the word gl_{q-1} with the label s_{q-1} along with gl_q with a label s_q. Dynamic programming can be a promising solution to amplify the log-likelihood of input training information set GL. To get the ideal label sequence for a specific input phrase, the Viterbi algorithm can be used [35].

5 Sanitization

The proposed δ_r-sanitizer is an enhanced sanitization model that addresses the problems of conventional IC-based sanitization [24]. The result shows that the proposed model is capable of sanitizing sensitive information more accurately. It reduces the issues associated with ambiguous language, the non-logical nature of the global IC computation, extra sanitation (which results in the least data utilization), and minimizes the computation cost. We used the SNOMED-CT to sanitize the previously detected sensitive terms (*SenT*) by replacing them with their generalized *Sanitized*(*SenT*) terms. Every sensitive term ($SenT_i$) should be safeguarded, so privacy is not disclosed to the attacker. The threshold is not required in the proposed model (e.g., the threshold, *beta*, is utilized in [17]). Conventionally, the threshold is set according to the least IC available among the terms, which may over-fit the sanitizing model and result in degraded utility [25]. We introduced δ as a limen for sanitation, influenced by K-anonymity, L-diversity, T-closeness, Differential Privacy (DP), and Relax form of DP [9]. These are popular and well-practised models used for privacy-preserving of structured data-sets. The δ is adjustable by the data-holders for the protection of the information as per their requirement and according to the sensitivity of each medical datasets. Equation 3 is used to set the threshold to user-defined δ.

$$Sanitized_{\forall SenT_i \in SenT} = C\_T_r(SenT_i) \tag{3}$$

where rand$(1, \delta)$, $\delta \leq$ gen and $C\_T = C\_T_1, C\_T_2, \dots, C\_T_{gen}$. In Eq. 3, we proposed the basic formulation for δ as a threshold. If the value of δ is less than the length of its generalized hierarchy (C_T), all those sensitive terms $SenT_i \in SenT$ are generalized/Sanitized randomly up to the user-defined δ level.

$$Sanitized_{\forall SenT_i \in SenT} = S_n(SenT_i) \tag{4}$$

where $\delta >$ gen. In Eq. 4, we show that if the value of $\delta > gen$ (length of its generalized hierarchy or clinical taxonomy), then all those sensitive terms $SenT_i \in SenT$ are replaced with their most generalized terms, which is C_T_{gen}. From the set of generalized/sanitized terms *sanitize*($SenT_i$) for each sensitive term $SenT_i$, the generalized term that fulfills the conditions of Eq. 5 is selected as optimal generalized term.

$$\begin{aligned} &Sanitized(SenT_i) = ((SenT_i \in C\_T_{i-gen}) \\ &\wedge (C\_T_r(SenT_i), r(l,\delta), \delta = 1 - gen, \\ &if\,\delta \le gen) \vee (C\_T_{gen}(SenT_i), \delta = n,\ if\ \delta > gen)) \end{aligned} \tag{5}$$

Equation (5) is a combination of Eqs. 3 and 4. Based on the afore-mentioned equations, the privacy bounds of proposed sanitization is defined by the following lemmas, which limit the disclosure risk of an adversary:

Proposition 1 *If the privacy threshold δ is less than n, then the generalization/sanitization level of $SenT_i$ is randomly selected between* 1 *to δ, where the probability of threshold selection is $P = 1/\delta$.*

Proof The generalization of $SenT_i$ is 1 to δ. In contrast if it is $\delta + 1$ to n, the probability of δ is always greater than 1. This contradicts the rule where the sum of all probability should equal to 1. Hence, the generalization threshold cannot be $\delta + 1$ to n. □

Proposition 2 *If δ is greater than C_T_n, then $SenT_i$ is a value of most generalized term in hierarchy C_T_{gen} against a sensitive term, where the probability of threshold selection is $P = 1/C\_T_n$.*

Proof The C_T_n is the total number of generalized terms for *i*th sensitive term. A threshold δ value greater than C_T_n refers to the selection of a generalized term that does not exist. Therefore, the generalized term should be selected between 1 and C_T_n with $P = 1/C\_T_n$ threshold probability. □

Algorithm 1 Sensitive terms sanitizer (δ_r).

```
Require: $SenT, MKB, \delta$
$\delta_r - Sanitizar(SenT, MKB, \delta)$
{
for each $SenT_i \in SenT$ do
    if $SenT_i \in MKBs$ then
        $C\_T = Get\_Clinical\_Taxonomy(SenT_i, MKB)$
        $C\_T_n = length(C\_T)$
        if $\delta \le C\_T_n$ then
            $r = rand(1, \delta)$
            $sanitize(SenT_i) = C\_T_r$
        else
            $sanitize(SenT_i) = C\_T_{gen}$
        end if
    else
        $sanitize(SenT_i) \leftarrow C\_T_{gen}$
    end if
    $sanitized_i \leftarrow sanitize(SenT_i)$
end for
return $sanitized_{list}$
}
```

The entire functionality of the proposed δ_r-Sanitizer is demonstrated in Algorithm 1. We have previously published a detailed formal verification and complexity analysis of the proposed sanitization algorithm in [36]. The δ_r-Sanitizer is fed by the sensitive medical entities (*SenT*). For the extraction of generalised medical entities, the medical knowledge-based data (*MKB*) is provided as input together with a user-defined threshold (δ_r). It produces a list of generalised words for each input parameter. Algorithm Line 1 specifies the inputs,i.e., knowledge-based terms (in this instance, SNOMED-CT), a list of sensitive words, and the privacy threshold δ. Then it is determined whether each sensitive term $SenT_i$ is defined in MKB in Lines 2 and 3. It obtains its full hierarchy of generalisations at line 5. It also obtains the length of the generalised hierarchy for each $SenT_i$ on line 6. On Line 7, it is checked whether the threshold δ is less than or equal to the hierarchy's length C_T or not. It retrieves generalized words between 1 and δ at random, in lines 8 and 9. It is generalized to as "finding" otherwise, on line 11. But if the line 4 condition is false, the most generalised word, in line 14, is used. Line 16 stores each generalised term one after the other, and line 18 returns the whole list to the caller model. The suggested method is more cost-effective, having an $O(n)$ time complexity than [17], which has a $O(n2)$ time complexity. For each sensitive word based on the available domain hierarchy, the proposed model increases security by randomly generating the value of δ. As a result, it is almost difficult for an opponent to deduce the sensitive term from the generalised ones.

Proposition 3 *The probability of finding* δ *is* $P(\delta) = 1/C\_T$ *for a single term, and* $P(\delta(SenT_i)) = 1/C\_T_n$ *for all terms. Therefore,*

$$P_{SenT} = \prod_{i=i}^{C\_T_n} P(\delta(SenT_i))$$

.

Proof The probability of disclosure risk is always less than or equal to one. Therefore, for each $P(C\_T_n) < 1$, the combined probability of all terms P_{SenT} is between 0 to 1 and starts approaching zero as the number of terms increases. □

Example For instance, $P(\delta_1) = 0.1$ and $P(\delta_2) = 0.2$, then, the $P_{SenT} = 0.02$. Similarly, $P(\delta_1) = 0.1$, $P(\delta_2) = 0.2$, $P(\delta_3) = 0.3$, and $P(\delta_4) = 0.05$, then, the $P_{SenT} = 0.0003$. It show, when we increase the number of sensitive terms, the probability of revealing original terms decrease toward to zero. This exemplifies that if there are more sensitive terms, in a document, the probability of original term disclosure will decrease and making it hard for an adversary to successfully breach the confidentiality of the sensitive data.

6 Performance Evaluation

This section details the Simulation Settings of the experimentation.

6.1 Configuration Settings

In simulations, we utilize the n2c2-2010 data-set [37] aimed at extracting medical concepts, e.g., health reports, medication lists, and test results. The evaluation is conducted by adopting the following performance metrics:

6.1.1 Evaluation Matrices

To estimated clinical corpus, precision, recall, and F1-score are used as evaluation matrices. The explanation of each evaluation criteria is given below:

- Precision: The Precision (AKA specificity) is the total true positive proportion to the whole of the true positive (*tp*) and false positive (*fp*) instances. To increase precision, the true positive rate is maximized, and the false positive should be minimized. The false-positive rate is devastating in the medical field. The proposed model aims to achieve the highest precision ratio. The formula to calculate precision is given below: $precision = tp/tp + fp$
- Recall: The recall (aka sensitivity) is the proportion of true positive instances to the whole of true positive and false-negative rates. To increase the recall, the true positive rate is maximized, and the false-negative rate is minimized. The formula for the recall is given below: $recall = tp/tp + fn$
- F1-score: F1-score takes effect (harmonic mean) of both the precision and the recall. The more the F1-score, the more the precision and recall. The value of the F1-score is usually biased towards the lower notwithstanding of recall or precision. The formula to evaluate F-1 is as follows: $F - 1 = 2 * precision * recall/Precision + recall$

To build the recognition model for unstructured medical data, the Python built-in library called keras [39] is utilized. These three measures before preparation are followed to get accurate results:

(1) Reducing noise: The stop words, Punctuation, and white spaces are cleared.
(2) Sentences Padding: The sentence padding (i.e., 250) is used to make input phrase size alike.
(3) Normalization: The corpora is turned to lower cases to get the word normalization.

A pre-trained Word2Vec model expresses the token level illustration. The default token embedding size is used as set out in the pre-trained Bio-Word2Vec model. The character embedding size is randomly defined between the −1 to 1 range and trained with Bi-LSTM. Four hundred dimensions for word embedding and 200 dimensions for character embedding are used. The dimension size for token embedding is 250, and for the character embedding, 150 are not changeable. Both embeddings are joined, resultant as 600-dimensional size, and transferred to the following contextual layer. The three CNN models with three window sizes (i.e., 3, 5, and 7) are trained in the contextual layer. The regional contextual illustration of every CNN design attained 200 (transformed from 600 to 200) dimensions as input by combined embedding. This illustration of every local contextual CNN design is combined again with 600 dimensions and transferred to Bi-directional LSTM to generate worldwide content. Then, the regional and worldwide content is transferred to CRF to estimate the labels. In the final step, the BIO (Beginning of entity, Intermediate of an entity, and end of an entity) tagging scheme [35] and the CRF are combined for sequence labeling.

Table 3 Comparision of propsed MER model with current state-of-the-art models

Methods	Recall	Precision	F-1 score
Glov+ LSTM+CRF [38]	83.6	84.0	83.8
Word2Vec+ LSTM+CRF [35]	–	–	85.8
ElMo+ LSTM+CRF [20]	86.2	87.4	86.8
Bio-BERT+ LSTM+Inference layer [22]	85.4	87.5	86.4
BERT+ LSTM+Inference layer [21]	–	–	89.5
Proposed (CNN-BiLSTM+CRF)	91.14	92.63	92.0

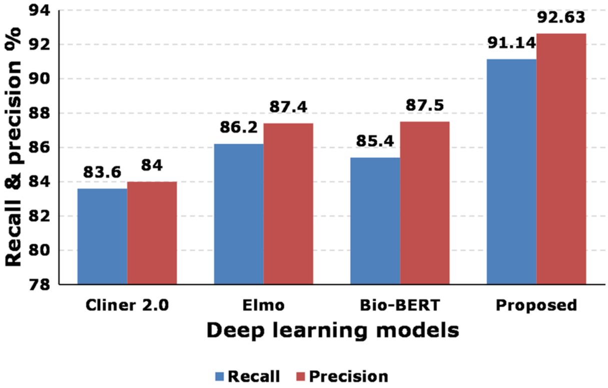


Fig. 5 Comparison of recall and precision among proposed MER and existing model

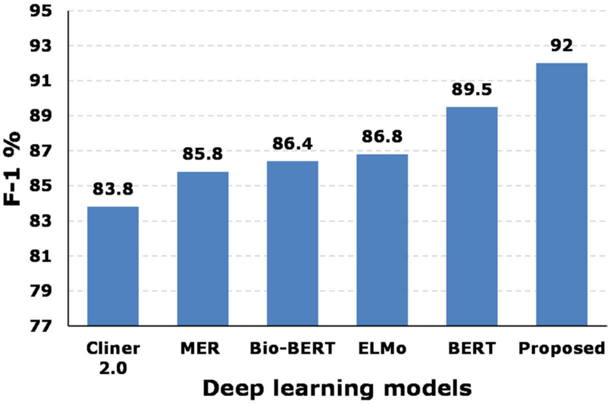


Fig. 6 Comparison of F1-score among proposed MER and existing model

6.2 Experimental Results

We now discuss the obtained numerical results:

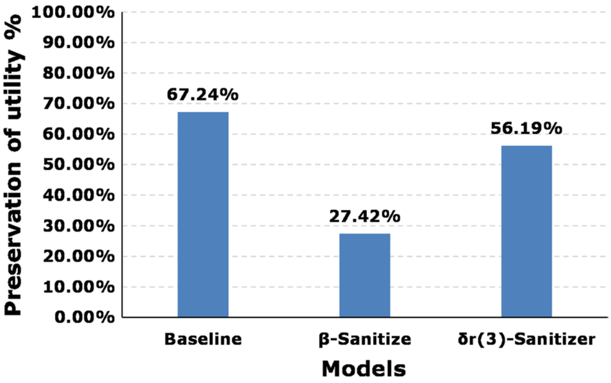


Fig. 7 Comparison of utility preservation using baseline sanitization, β-based sanitizer, and δ_r-sanitizer

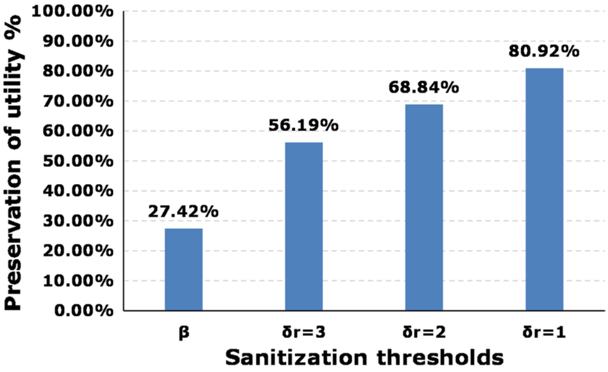


Fig. 8 Comparison of different variants of δ_r-sanitizer with β-based sanitizer

6.2.1 MER Detection Accuracy

Table 3 presents the comparison among proposed MER and the up-to-date state-of-the-arts. We compared our model with [20, 21, 22, 35, 38]. Figure 5 illustrates that the recall and precision results. The figure shows that the recall value for the proposed model is 91.14% with 92.63% precision. Whereas, recall value for [20, 38], and for [22] is 83.6 percent, 86.2 percent, and 85.5 percent, respectively. Similarly, the precision percentage is 84 percent, 87.4 percent, and 87.5% for the said models, respectively. The F1-score is depicted in Fig. 6, which illustrates that for the proposed model is 92 percent. whereas, it is 83.8, 85.8, 86.4, 86.8, and 89.5

6.2.2 δ_r-Sanitizer Utility Preservation

Figure 7 presents the comparison of utility preservation among the proposed model, [24, 25], and with the baseline provided by Beaumont Hospital, Dublin, Ireland. The utility preservation in the proposed model is 56.19 percent, for β-based models, it is 27.42, and for the baseline, it is 67.24 percent. The result shows that compared to the β-based models, the proposed model preserved more utility. However, the proposed utility preservation is 11.05% less than the baseline given by the hospital. Whereas, there is a 39.82% difference in baseline and β-based models. In the provided baseline, the threshold value is least, whereas, in the proposed model, we set the threshold value as 3. Figure 8 represents different variants of δ_r-sanitizer with respect to different δ values. It also shows the utility preservation comparison of these variants with β-based sanitizer. The utility preservation percentage for β-based sanitizer is 27.42 percent. Whereas, It changes with the changing value of δ in the proposed mechanism. It is 56.19, 68.84, and 80.92% for the δ_r-sanitizer, with δ equal to 3, 2, and 1. It is obvious that the value of δ influences the utility preservation of a model. That is why we made it open for the data-holder to set.

6.2.3 Latency

We compared latency between fog and cloud-based smart healthcare systems. Figure 9 illustrates that the response time of a fog-based application is less than that of a cloud-based application. The response time (in our case) may be defined as when a sensor device generates data and sends it to the upper layer (i.e., either to the fog or to the cloud), and response against that received data is triggered. As shown in the figure, we generate an equal number of messages sent to the fog and the cloud. The response time in fog-based applications is less than the cloud-based. For the first 5000 requests (messages), the time is 5 and 7 milliseconds for fog and cloud, respectively. We kept on increasing messages and noted time after every 5000 messages, as shown in the figure. For 100,000 messages, the response time of a fog-based application is 32 milliseconds, whereas, for the same number of messages, the cloud-based application took 36 milliseconds. On average, there is a 3-millisecond difference between them for all the responses.

7 Discussions

To better protect patients from the spread of infectious diseases, we developed a new deep learning-based sanitization method to identify and sanitise out clinical entities. The suggested model considers both global and local context, whereas previous research solely considered global context. We used three baseline systems as a benchmark to assess the proposed model (using i2b2-2010 data): (a) CRF model, (b) BLSTM model (i.e., token-level representation), and (c) BLSTM model (i.e., character- and token-level

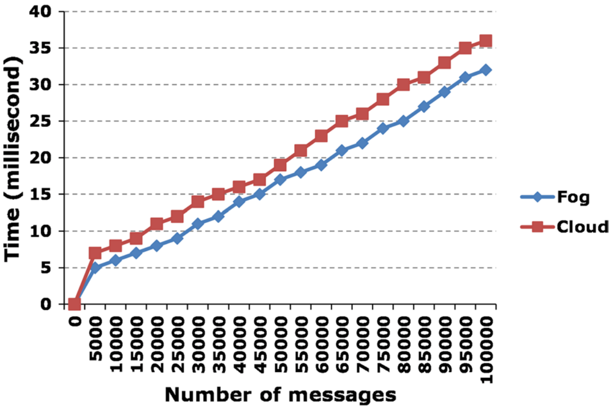


Fig. 9 Latency comparison between fog and cloud-based application

representation). We make comparisons to current models for a thorough review. While the CRF performs well in terms of Exact criteria with 84% of F1-score (as shown in Table 3), it still has a long list of steps, such as pre-processing and tuning of parameters to reach a satisfactory accuracy. The implementation of deep learning (i.e., CNN and RNN) models is the answer to the troubles of CRF. In the context of CRF, deep learning models can improve accuracy. For example, the LSTM-CRF model delivers an F1-score of 86% and 92% on Exact and Inexact evaluation criteria, respectively. The traditional approaches using contextual embedding provide an F1-score of 87.90%, whereas the basic embedding models have an F1-score of 84.86%. While standard embedding (i.e., GloVec and word2vec) seems to have a 2.3% advantage over the traditional approach in this instance, ELMo and BERT have a smaller advantage (i.e., about 1.5%). While spending extra computational resources and effort, advanced and sophisticated embedding boosts model performance by just 2.3%.

The LSTM and its many variants placed the most emphasis on the global context. They sometimes neglected to include the local context. The information in the medical records is all connected and has to be gathered. While the BERT embedding may help certain LSTM-based models, the overall quality of the findings is not significant. The new context considers both the local and global context and aims to identify the clinical entities. The BiLSTM model captures global context, whereas the CNN model captures local context. Higher accuracy may be achieved with BERT with advanced embedding by investing in additional computer resources, yielding a 1.2% improvement in accuracy. In comparison, our model is better, achieving 4.5% greater outcomes than current state-of-the-art algorithms. Compared to the CRF-based model, the method presented in this paper yields an 8% improvement in accuracy. The information in this study enables the incorporation of the local context, which is used to identify the biological entities. Additionally, in the sanitization phase, Information Content (IC)-based approaches often result in incorrect outcomes, as noted in [17]. The suggested sanitization strategy surmounts the limitations of IC-based procedures, as noted in [17, 23, 25]. Moreover, the IC of "*Virus*" cannot differentiate between a computer and biological viruses. It gives confusing results since it overestimates the IC value. Further, in these experiments, a fixed privacy threshold (β) was utilised. Nevertheless, the trade-off between privacy and data usefulness is an individual one. The individual's privacy threshold should be adjustable to accommodate their individual needs.

8 Conclusions and Future Work

The introduction of fog computing in smart healthcare infrastructures provides a promising solution at the edge of the network when it comes to latency. It may help advance the present medical research and diagnose various diseases or find the solution to future medical field challenges. It is important to apply privacy-preserving methods to secure the patients' confidential clinical information before providing it for research purposes. However, present studies show privacy violations and less data utilization due to improper handling of utilization and privacy issues. The mechanism used fog-enabled deep learning models and a novel sanitation mechanism for preserving the privacy and utility of medical entities concerning data-holder needs. The Deep Privacy framework improved detection accuracy by 92% in comparison to other deep learning models. Improvement of 56.19%

had been seen in the utility rate of sanitized medical documents with a value of δ equal to 3 with reduced latency.

Although the presented approach provides flexible privacy protection via user-defined privacy thresholds, it falls short of providing full semantic privacy protection, such as differential privacy. Thus, one potential future study path is to enhance the suggested solution's semantic privacy assurance and mathematical verification. Additionally, the suggested approach sanitizes the sensitive words only. Another potential study area is the identification and sanitization of semantically related words.

Author Contributions SAM (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Software: Lead; Validation: Lead; Writing -original draft: Lead) NT (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Software: Lead; Validation: Lead; Writing - original draft: Lead) AA (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Software: Lead; Validation: Lead; Writing - original draft: Lead) AA (Conceptualization: Supporting; Data curation: Supporting; Writing - original draft: Supporting) SURM (Conceptualization: Lead; Data curation: Lead; Formal analysis: Lead; Funding acquisition: Lead; Investigation: Lead; Methodology: Lead; Software: Lead; Validation: Lead; Writing - original draft: Lead) HM (Conceptualization: Supporting; Data curation: Supporting; Writing - original draft: Supporting) Haris Pervaiz (Conceptualization: Supporting; Data curation: Supporting; Writing - original draft: Supporting) SSG (Supervision: Supporting; Writing - review & editing: Supporting)

Funding This work was supported in part by the National Natural Science Foundation of China (NSFC) Project No. 61950410603.

Data Availability Data and material is available here NA.

Code Availability Software application code is available here NA.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Ning, Z., Dong, P., Wang, X., Hu, X., Guo, L., Hu, B., Guo, Y., Qiu, T., & Kwok, R. (2020). Mobile edge computing enabled 5g health monitoring for internet of medical things: A decentralized game theoretic approach. *IEEE Journal on Selected Areas in Communications, 39*, 463–478.
2. Liao, H., Zhou, Z., Zhao, X., Zhang, L., Mumtaz, S., Jolfaei, A., Ahmed, S. H., & Bashir, A. K. (2020). Learning-based context-aware resource allocation for edge-computing-empowered industrial IoT. *IEEE Internet of Things Journal, 7*(5), 4260–4277.
3. Petropoulos, A., Sikeridis, D., & Antonakopoulos, T. (2020). Wearable smart health advisors: An imu-enabled posture monitor. *IEEE Consumer Electronics Magazine*.
4. Rahmani, A. M., Gia, T. N., Negash, B., Anzanpour, A., Azimi, I., Jiang, M., & Liljeberg, P. (2018). Exploiting smart e-health gateways at the edge of healthcare internet-of-things: A fog computing approach. *Future Generation Computer Systems, 78*, 641–658.
5. Tuli, S., Tuli, S., Wander, G., Wander, P., Gill, S. S., Dustdar, S., et al. (2020). Next generation technologies for smart healthcare: challenges, vision, model, trends and future directions. *Internet Technology Letters, 3*(2), e145.
6. Ud Din, I., Guizani, M., Hassan, S., Kim, B., Khurram Khan, M., Atiquzzaman, M., & Ahmed, S. H. (2019). The internet of things: A review of enabled technologies and future challenges. *IEEE Access, 7*, 7606–7640.

7. Gill, S. S., Xu, M., Ottaviani, C., Patros, P., Bahsoon, R., Shaghaghi, R., et al. (2022). AI for next generation computing: Emerging trends and future directions. *Internet of Things, 19*, 100514.
8. Tariq, N., Asim, M., Al-Obeidat, F., Zubair Farooqi, M., Baker, T., Hammoudeh, M., & Ghafir, I. (2019). The security of big data in fog-enabled IoT applications including blockchain: a survey. *Sensors, 19*(8), 1788.
9. Moqurrab, S. A., Anjum, A., Manzoor, U., Nefti, S., Ahmad, N., & Ur Rehman Malik, S. (2017). Differential average diversity: An efficient privacy mechanism for electronic health records. *Journal of Medical Imaging and Health Informatics, 7*(6), 1177–1187.
10. Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*
11. Moqurrab, A., Ayub, U., Anjum, A., Asghar, S., & Srivastava, G. (2021). An accurate deep learning model for clinical entity recognition from clinical notes. *IEEE Journal of Biomedical and Health Informatics*.
12. Ma, J., Huang, X., Mu, Y., & Deng, R. H. (2020). Authenticated data redaction with accountability and transparency. *IEEE Transactions on Dependable and Secure Computing*
13. Tariq, N., Khan, F. A., & Asim, M. (2021). Security challenges and requirements for smart internet of things applications: A comprehensive analysis. *Procedia Computer Science, 191*, 425–430.
14. Tariq, N., Asim, M., Khan, F. A., Baker, T., Khalid, U., & Derhab, A. (2021). A blockchain-based multi-mobile code-driven trust mechanism for detecting internal attacks in internet of things. *Sensors, 21*(1), 23.
15. Shukla, S., Thakur, S., Hussain, S., Breslin, J. G., & Jameel, S. M. (2021). Identification and authentication in healthcare internet-of-things using integrated fog computing based blockchain model. *Internet of Things, 15*, 100422.
16. Buyya, R. H., Calheiros, R. N., & Dastjerdi, A. V. (2016). *Big Data: Principles and Paradigms*. Morgan Kaufmann.
17. Iwendi, C., Moqurrab, S. A., Anjum, A., Khan, S., Mohan, S., & Srivastava, G. (2020). N-sanitization: A semantic privacy-preserving framework for unstructured medical datasets. *Computer Communications*.
18. Habibi, M., Weber, L., Neves, M., Wiegandt, D. L., & Leser, U. (2017). Deep learning with word embeddings improves biomedical named entity recognition. *Bioinformatics, 33*(14), i37–i48.
19. Unanue, I. J., Borzeshi, E. Z., & Piccardi, M. (2017). Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition. *Journal of Biomedical Informatics, 76*, 102–109.
20. Zhu, H., Paschalidis, I. C., & Tahmasebi, A. (2018). Clinical concept extraction with contextual word embedding. arXiv preprint arXiv:1810.10566.
21. Si, Y., Wang, J., Xu, H., & Roberts, K. (2019). Enhancing clinical concept extraction with contextual embeddings. *Journal of the American Medical Informatics Association, 26*(11), 1297–1304.
22. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). Biobert: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics, 36*(4), 1234–1240.
23. Batet, M., & Sánchez, D. (2014). Privacy protection of textual medical documents. In: IEEE Network Operations and Management Symposium (NOMS). IEEE, pp. 1–6.
24. Sanchez, D., & Batet, M. (2017). Toward sensitive document release with privacy guarantees. *Engineering Applications of Artificial Intelligence, 59*, 23–34.
25. Batet, M., & Sánchez, D. (2019). Leveraging synonymy and polysemy to improve semantic similarity assessments based on intrinsic information content. *Artificial Intelligence Review, 53*, 2023–2041.
26. Saha, R., Kumar, G., Rai, M. K., Thomas, R., & Lim, S.-J. (2019). Privacy ensured *e*-healthcare for fog-enhanced IoT based applications. *IEEE Access, 7*, 44536–44543.
27. Zhao, O., Liu, X., Li, X., Singh, P., & Wu, F. (2020). Privacy-preserving data aggregation scheme for edge computing supported vehicular ad hoc networks. *Transactions on Emerging Telecommunications Technologies*, p. e3952.
28. Dong, P., Ning, Z., Obaidat, M. S., Jiang, X., Guo, Y., Hu, X., Hu, B., & Sadoun, B. (2020). Edge computing based healthcare systems: Enabling decentralized health monitoring in internet of medical things. *IEEE Network, 34*, 254–261.
29. Cui, J., Wei, L., Zhong, H., Zhang, J., Xu, Y., & Liu, L. (2020). Edge computing in vanets: An efficient and privacy-preserving cooperative downloading scheme. *IEEE Journal on Selected Areas in Communications, 38*(6), 1191–1204.
30. Wang, X., Feng, Y., Ning, Z., Hu, X., Kong, X., Hu, B., & Guo, Y. (2020). A collective filtering based content transmission scheme in edge of vehicles. *Information Sciences, 506*, 161–173.

31. Bouchelaghem, S., & Omar, M. (2020). Secure and efficient pseudonymization for privacy-preserving vehicular communications in smart cities. *Computers and Electrical Engineering, 82*, 106557.
32. Guan, Z., Zhang, Y., Wu, L., Wu, J., Li, J., Ma, Y., & Hu, J. (2019). Appa: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT. *Journal of Network and Computer Applications, 125*, 82–92.
33. Tariq, N., Asim, M., Maamar, Z., Farooqi, M. Z., Faci, N., & Baker, T. (2019). A mobile code-driven trust mechanism for detecting internal attacks in sensor node-powered IoT. *Journal of Parallel and Distributed Computing, 134*, 198–206.
34. Alahmar, A. D., & Benlamri, R. (2020). Snomed ct-based standardized e-clinical pathways for enabling big data analytics in healthcare. *IEEE Access, 8*, 92765–92775.
35. Liu, Z., Yang, M., Wang, X., Chen, Q., Tang, B., Wang, Z., & Xu, H. (2017). Entity recognition from clinical texts via recurrent neural network. *BMC Medical Informatics and Decision Making, 17*(2), 67.
36. Kanwal, T., Moqurrab, S. A., Anjum, A., Khan, A., Rodrigues, J. J., & Jeon, G. (2021). Formal verification and complexity analysis of confidentiality aware textual clinical documents framework. *International Journal of Intelligent Systems*
37. Uzuner, Ö., South, B. R., Shen, S., & DuVall, S. L. (2011). 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association, 18*(5), 552–556.
38. Boag, W., Sergeeva, E., Kulshreshtha, S., Szolovits, P., Rumshisky, A., & Naumann, T. (2018). Cliner 2.0: Accessible and accurate clinical concept extraction. arXiv preprint arXiv:1803.02245.
39. Chollet, F. et al. (2015). Keras. https://github.com/fchollet/keras.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Syed Atif Moqurrab is Assistant Professor at Air University, Pakistan. He is pursuing PhD from COMSATS, Pakistan and has received MS (Computer Science) FAST-NU, Pakistan. His research interests include Data Privacy, Artificial Intelligence, Digital Marketing and full stack Development.

Noshina Tariq received the M.S. and Ph.D. degrees in computer science from the Department of Computer Science, National University of Computer and Emerging Sciences, Islamabad, Pakistan. She is currently working as an Assistant Professor with the Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), Islamabad. Her research interests include the Internet of Things, fog computing, cyber security, blockchain, and machine learning.

Adeel Anjum is an Assistant Professor in the department of Computer Sciences at CIIT Islamabad Campus. He completed his PhD with distinction in the year 2013. His area of research is data privacy using artificial intelligence techniques. He has several publications in international conferences. He is also the author of a book on data privacy. He serves in the technical program committees of various international conferences.

Alia Asheralieva obtained her Ph.D. from the University of Newcastle, Australia, in 2014. In 2015 and 2016, she was with the Graduate School of Information Science and Technology, Hokkaido University, Japan. From 2017, she was with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design. She is currently with the Department of Computer Science and Engineering, Southern University of Science and Technology, China. Her main research interests span many areas of communications and networking, including cloud/edge/ fog computing, blockchains, device-to-device and Internet of Things communications, game theory, machine learning, stochastic and combinatorial optimization.

Saif U. R. Malik received the Ph.D. degree from North Dakota State University, USA, in 2014. He is currently working as a Senior Researcher with Cybernetica AS, Estonia. Previously, he worked as an Assistant Professor with the Department of Computer Science, COMSATS University Islamabad, Pakistan. He is also an Active Researcher in the field of cloud computing, data centers, formal methods and its application in large scale computing systems, and data security and privacy. His work has appeared in highly reputable venues.

Hassan Malik (Member, IEEE) received the B.E. degree from the National University of Sciences and Technology (NUST), Pakistan, in 2009, the M.Sc. degree from the University of Oulu, Finland, in 2012, and the Ph.D. degree from the 5G Innovation Centre (5GIC), University of Surrey, U.K., in 2017. He worked as a Research Assistant with the Centre for Wireless Communication (CWC), Oulu, Finland, from 2011 to 2013. From 2017 to 2020, he worked as a Researcher with the Thomas Johann Seebeck Department of Electronics, Tallinn University of Technology, Estonia. He is currently working as a Senior Lecturer with the Computer Science Department, Edge Hill University, U.K. His current research interests include wireless communication protocols, the IoT, URLLC, LPWAN technologies, blockchain-based IoT, and embedded systems.

Haris Pervaiz (Member, IEEE) received the M.Sc. degree in information security from the Royal Holloway University of London, Egham, U.K., in 2005, and the Ph.D. degree from the School of Computing and Communication, Lancaster University, Lancaster, U.K., in 2016. He was a Research Fellow with the 5G Innovation Centre, University of Surrey, Guildford, U.K., from 2017 to 2018, and an EPSRC Doctoral Prize Fellow with the School of Computing and Communication, Lancaster University, from 2016 to 2017. He is currently working as a Lecturer with the InfoLab21, Lancaster University. He has been actively involved in projects, such as CROWN, CogGREEN, TWEETHER, and Energy Proportional EnodeB for LTE-Advanced and Beyond and the DARE project, and an ESPRC funded project. His current research interests include green heterogeneous wireless communications and networking, 5G and beyond, millimeter wave communication, and energy and spectral efficiency. He is also an Associate Editor of IEEE ACCESS, IET Quantum Communications, IET Network, Emerging Telecommunications Technologies (Wiley), and Internet Technology Letters (Wiley).

Sukhpal Singh Gill is a Lecturer (Assistant Professor) in Cloud Computing at School of Electronic Engineering and Computer Science, Queen Mary University of London, UK. Prior to this, Dr. Gill has held positions as a Research Associate at the School of Computing and Communications, Lancaster University, UK and also as a Postdoctoral Research Fellow at CLOUDS Laboratory, The University of Melbourne, Australia. Dr. Gill is serving as an Associate Editor in Wiley ETT and IET Networks Journal. His research interests include Cloud Computing, Fog Computing, Software Engineering, Internet of Things and Healthcare. For further information, please visit http://www.ssgill.me

Authors and Affiliations

Syed Atif Moqurrab[1] · Noshina Tariq[2] · Adeel Anjum[1,3] · Alia Asheralieva[3] · Saif U. R. Malik[4] · Hassan Malik[5] · Haris Pervaiz[6] · Sukhpal Singh Gill[7]

Syed Atif Moqurrab
atifmoqurrab@gmail.com

Noshina Tariq
dr.noshina@szabist-isb.edu.pk

Adeel Anjum
adeel.anjum@comsats.edu.pk

Alia Asheralieva
asheralievaa@sustech.edu.cn

Saif U. R. Malik
saif.rehmanmalik@cyber.ee

Hassan Malik
Malikh@edgehill.ac.uk

Haris Pervaiz
h.b.pervaiz@lancaster.ac.uk

[1] Department of Computer Sciences, COMSATS University, Islamabad, Pakistan

[2] Department of Computer Science, Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan

[3] Department of Computer Science and Engineering, Southern University of Science and Technology, Nanshan District, Shenzhen, Guangdong, China

[4] Cybernetica AS Estonia, Tallinn, Estonia

[5] Department of Computer Science, Edge Hill University, Ormskirk, UK

[6] School of Computing and Communications, Lancaster University, Lancashire, UK

[7] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK