International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing K-Clustering based Privacy Preserving for E-Healthcare IoT Systems

P. Umaeswari[1], G. Gayathiri Devi[2], Gomathi. C[3], Anusuri Krishna Veni[4], K. B. Kishore Mohan[5]

Submitted: 10/12/2023 **Revised:** 22/01/2024 **Accepted:** 01/02/2024

Abstract: New technologies in the cloud, social media, the Internet of things (IoT), and E-healthcare systems all need privacy protection. Health and medical records, in terms of strategy, include images and health records about patients, and all personal details must be kept private to protect patients' privacy. Standard encryption systems for textual and structural one-dimensional data could not be specifically extended to e-health data due to weaknesses of digital data structures. When medical data is exchanged in the general public for various study purposes, an effective data retention system with minimum information loss is needed. This research has suggested a medical healthcare IoT's-driven infrastructure with limited access based on the inspiration. Three techniques are used in the infrastructure. The first technique preserves a patient's sensitive information by quantifying the least amount of information lost during the anonymization phase. Based on the clustering principle, the technique has also designed data access that provides the public, doctors, and nurses access to a patient's sensitivity information. K-anonymity privacy protection depends on local encryption, which is based on cell suppression, which is the second suggested technique. This approach employs a mapping approach to divide the data into various regions in such a way that data from the same population is grouped. Finally, data processing techniques (such as k-means) are often used to filter data obtained from wireless sensor networks to make medical recommendations to doctors and patients. Many approaches, on the other hand, face a risk of data loss during the data handling process. Extensive simulations are used to compare the proposed algorithm's efficiency to that of the state-of-the-art algorithm. Simulation findings show that the suggested algorithms are beneficial in terms of an effective cluster forming in a small amount of time, minimal data loss, and data propagation execution time.

Keywords: *Privacy-Preserving; e-Healthcare; Internet of Things (IoT); Clustering; K-means; K-medoid; K-Anonymity*

1. Introduction

The emergence of the Internet of Things (IoT) [1] allows any entity to be transformed into network data via wireless sensor networks. This has resulted in a wide range of technologies, including smart cities, smart houses, and smart grids. Furthermore, leading to its efficient capabilities of incorporating within infrastructure resources and provisioning critical knowledge to patients, IoT has been broadly implemented to design and facilitate healthcare networks. Wireless Sensor Networks (WSNs) can gather a wide range of data to aid the spread of many e-health programs, such as electronic health cards, online patient control, and health portals. E-Healthcare IoT system [2] has demonstrated great promise in improving people's health and can be used for a variety of uses, from implantable medical implants to Wireless Body Area Networks (WBAN).

Generally, the E-Healthcare IoT system is consists of 3 layers: the perception layer, the network layer, and the application layer, as shown in Figure 1.

Perception Layer: The basis of Internet of Things e-healthcare solutions is the perception layer. It provides the flexible attachment of numerous sensor devices through WSNs. Wearable sensors and implant sensors are two types of sensors used in e-healthcare systems. These sensors can collect and transmit critical healthcare data from patients to the network layer.

Network Layer: In a healthcare IoT scheme, the network layer serves as an auxiliary bridge. It is primarily in charge of encryption, authentication, and dissemination of confidential healthcare information.

Platform Layer: The platform layer is responsible for providing valuable IoT software. Through collecting and reviewing the gathered healthcare data, mainly integrates intelligent decision making and service provision.

Security and privacy of data pertaining to patients are two crucial concepts. While data privacy involves the data being available to those who are authorized to view and use it, data security relates to the storage and transmission of data in a secured way to assure its legitimacy, authenticity, and secrecy [3]. Effective defence mechanisms could be developed by utilizing different approaches and standards. People's wellbeing is increased by the widespread

[1]*Associate Professor, Department of Computer Science and Business Systems, R.M.K. Engineering College, Chennai.*
Email: umaeswari11@gmail.com
ORCID: 0000-0003-2870-6403s
[2]*Associate Professor, Department of Science and Humanities, R.M.D. Engineering College, Kavaraipettai, Thiruvallur District, Tamil Nadu, India.*
Email: gayathiri77@gmail.com
ORCID: 0000-0002-7581-1815
[3]*Assistant Professor. Department of AI & DS, Panimalar Engineering College, Chennai.*
Email: gomathipec@gmail.com
ORCID: 0009-0003-9938-7013
[4]*Assistant Professor, Department of Computer Science & Engineering - Data Science, Madanapalle Institute of Technology & Science, Kadiri Road, Angallu, Madanapalle- 517325, Andhra Pradesh, India.*
Email: krishnavenianusuri35@gmail.com
ORCID: 0009-0003-6129-0910
[5]*Associate Professor, Department of Bio Medical Engineering, Saveetha Engineering College, Chennai.*
Email: kishoremohan@saveetha.ac.in
ORCID: 0009-0002-6305-1457

use of IoT-enabled e-healthcare equipment [4], but privacy and information security are also significantly burdened.

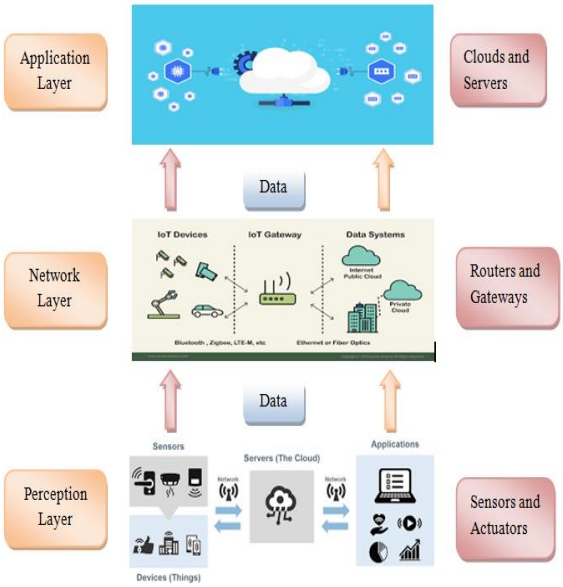

Fig. 1. E-healthcare IoT systems Framework

For cluster creation, the K-medoid ML technique is used, and the local recording method is used for data protection [5]. There is less risk of information leakage during data transfer where local recoding is used. Data privacy algorithms focused on clustering as K users [6] have also been introduced in previous studies. However, it has been pointed out that these algorithms take a long time to form clusters and suffer from knowledge loss. For data protection, a major study has used anonymity, data masking, and data stuffing [7]. In the latest privacy algorithms, these challenges are NP-hard questions, and most of them have included challenges like "incognito, clustering, global recoding and variation of privacy". Furthermore, in the healthcare industry, cluster processes and local coding approaches have a greater effect on data protection [8]. The most critical factor is the protection of personal information. In the event of a privacy breach, the intruder is free to use any information for any reason. As a result, since any physical person or object may be manipulated, this research has centered on the difficult methods of information protection when data is shared by IoT devices. A significant proportion of computers and sensors in an IoT environment are powered by batteries, with limited batteries and low computing energy consumption, with data security being the most important consideration. To propose an algorithm for identifying information loss during the anonymization process based on a clustering technique such as K medoid, to reduce information loss and protect personal privacy in the healthcare domain.

To build a K-anonymity framework based on cell suppression for privacy preservation using a local recording. The local recoding approach uses a mapping approach to divide data into various regions and group data of each kind together. As a result, the beauty of using local recoding is that it prevents third-party consumers, like physicians and nurses, from accessing actual patient data by inserting inaccurate values.

Through the analysis of the rapidly increasing sensor results, the k-means approach [9] has been commonly used to provide effective diagnostic suggestions for physicians and patients. Data processing positions are often outsourced to third-party entities of specialist consultants, given that healthcare systems are also challenged by restricted computational capacity. Patients' personal information could be traded to the dark market if an analyst were malicious, resulting in data privacy breaches and perhaps endangering patients' lives [10]. As a result, a widely researched subject has been how to interpret confidential data while maintaining participant privacy [11]. To avoid privacy leakage while using k-means algorithms, an interactive method [12] was created. However, since the cluster in this scheme is public, there is also a risk of privacy leakage, especially when participants communicate with an outsider who considers the cluster center, could effectively quantify the sensitive data of other cluster participants. Furthermore, for healthcare applications, an algorithm's time efficiency is critical [13]. Most modern privacy-preserving k-means algorithms are too time-consuming to preserve shared privacy.

This paper is organized as follows for the remainder of it. Section 2 presents relevant literature. The requirements for e-healthcare data security and privacy are covered in the third section. The fourth section presents the suggested K Clustering approaches. In the fifth section, the experimental results are discussed and the effectiveness of the suggested technique is tested. A conclusion is provided in Section 6.

2. Related Works

The study of k-means algorithms that preserve privacy has made a significant contribution. There have been several proposals for privacy-preserving k-means algorithms. "Vaidya et al. [14]", published the paper on privacy-preserving k means clustering techniques. They group vertical data thus lowering connectivity costs and ensuring that data protection is maintained properly. In the phase of measuring cluster centers, Jha et al. [15] found privacy preservation. They implemented two separate privacy protection systems; one depends on unaware polynomial analysis and the other on homomorphic encryption. When referring a participant to the nearest cluster center, however, the system did not take into account privacy concerns. "Bunn et al. [16]", suggested a two-party k-means clustering protocol in which the clustering results are calculated in each iteration instead of using the cluster center. During the clustering method, this protocol avoids cluster center leakage. Xing et al. [17] introduced a shared privacy defence method to stop the leak of private data and successfully defeat collusion risks.

The privacy defense of the k-means study has drawn a lot of interest in differential privacy. For the first time, Blum et al. [18] implemented a differential privacy function into a privacy-preserving approach. By applying noise to the cluster core, the proposed methodology lowers the chance of privacy leakage. The efficiency of the k-means clustering is, however, directly influenced by the initial center point collection. To allow for the random collection of the initial cluster core, Yang et al. [19] proposed an enhanced differential privacy (IDP) k-means technique. A unique blockchain-based data protection system [20] was developed, using blockchain technologies to eradicate the single-point-of-failure issue while still maintaining privacy through homomorphic encryption and differential privacy.

Nonlinear dynamics were presented in the literature in 2 ways: "discrete systems like logistic maps and continuous systems like hyper-chaotic systems" [21]. Because of their ability to generate random number sequences with excellent stochastic behavior,

these systems have been commonly used in cryptography. Furthermore, the chaotic maps' original values are highly vulnerable to any changes. These initial values were used as secret keys in chaos-cryptosystems based on this by cryptographers. The "chaotic-based pseudo-random number generator (PRNG)" methodology is one of the most well-known cryptographic implementations in this regard. A PRNG, which is similar to a random number generator, is commonly used to generate stream keys for messy encryption technology. Hu et al. [22] proposed a cryptographic PRNG based on a high-dimensional chaotic map containing a sequence of three chaotic orbits coordinates for image encryption and other cryptographic applications. High-dimensional chaotic maps [23] and, in general, lower-dimensional chaotic systems have been reported as problems with PRNGs in recent studies. Although the sequences created by high-dimensional chaotic systems cannot be predicted, they do have some drawbacks, such as poor performance and high computational complexity [24]. Randomness functions could be used to construct a series of real numbers using chaotic constructs. Despite this, the noisy map sequence does not correspond to the computations with the pixels in digital images perfectly and does not use a finite precision. Some solutions, like the encryption of particular images in real-time WCE [25] activities, are also unsuitable for immediate use and are not appropriate for real-time computing applications.

Luo et al. [26] have built on the foundations of [27] by improving the "secret sharing scheme (SW-SSS)", which optimizes the secret exchange and precisely repairs the shared data while storing the patient's data in the cloud. This research also argues that even if a computer is hacked, patient data is secure. The research [28] assessed and defined the model's privacy means, as well as the tradeoff between privacy, reliability, and consistency. The data protection issue triggered by a rise in operators and the legalization of IoT technology was studied by Zhou et al. [29]. Furthermore, a system was developed to examine and discover the implications of privacy and protection in the delivery of IoT emerging approaches. In paper [30] establishes a chaos-based encryption scheme for the security of patient records. Patient data is encrypted in the form of a casual picture, ensuring efficiency and the highest level of security against counter-attacks. For validation and privacy protection, the electrocardiogram (ECG) for IoT-based medical care was designed. To enhance the IoT ecosystem for data protection, the fog computing environment has been implemented for quick deployment and low latency. Even though fog computing has improved networking problems, data security in fog-based IoT implementations still needs to be successfully managed. In [31], the patient's revocation was introduced as a fundamental concept for applying blockchain technologies in healthcare. "Smart cross-domain data exchange, self-adaptive access control, and smart virtualization aids have been implemented for data security, restricted delivery, and user revocation".

The authors of [32] proposed the EETP-MAC protocol for transmitting patient data using prioritization by classifying into various perceptions while keeping energy usage in mind. [33] Presents an optimized architecture of MAC superframe structure for managing non-emergency data with improved efficiency. These reports, on the other hand, do not understand how to handle medical data from different environments. Mathematical research for cluster creation and knowledge sharing has validated the anonymization issue. As a result, existing findings on the data protection of patient health surveillance-related IoT distribution have prompted researchers to plan and improve reliable data sharing systems with high data privacy.

3. E-healthcare Data Protection and Privacy Standards

Including the fact that a large number of healthcare organizations may not assign sufficient resources to protection and privacy [34], security and privacy are improved considerably in E-healthcare IoT. This system generates a growing amount of complex real-time data that is highly sensitive. On one side, compromising the reliability of a medical device or network may have catastrophic repercussions. The patient's privacy knowledge, on the other hand, is present in the data acquisition, delivery, file storage, and republication processes. The basic standards mentioned below must be considered when creating medical Internet security and privacy methods [2]. Figure 2 shows the standard approaches of data encryption and decryption.

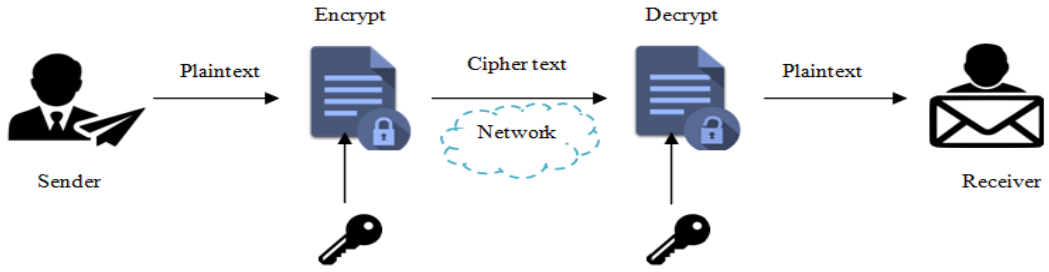

Fig. 2. Systematic Strategy of Data Encryption and decryption

- ***Data Integrity:*** The phrase of "data integrity" refers to the state in which every statistic qualities satisfy semantic expectations void of manipulation. It is reliable and precise to two degrees. International keys, restrictions, rules, and triggers can be used to protect four forms of data integrity: "entity integrity, domain integrity, referential integrity, and user-defined integrity".
- ***Data Usability:*** The term "Data Usability" relates to the ability of authorized users to access data and data structures. Although big data offers many benefits, it includes many disadvantages, such as contaminated and abnormal data. Furthermore, data corruption or destruction as a result of illegal access ruins data accessibility even more.
- ***Data Auditing:*** Auditing of the data Medical data access auditing is a standard measure for detecting and monitoring suspicious incidents, as well as an important way to control resource use. Moreover, cloud service providers frequently carry out problematic tasks, making the application of suitable auditing methods necessary. Common audit information includes users, cloud service providers, connection, and activity logs.

- ***Patient Data Privacy:*** Two categories exist for patient data: general documents and personal data. Sensitive information, sometimes referred to as medical privacy, includes things like mental health, sexual orientation, sexual functioning, infectious diseases, reproductive health, drug misuse, genetic records, and identifying information. It is imperative that we ensure that confidential data remains out of the hands of unauthorized individuals and that, in the event that it is retrieved, the information it contains is useless to them.

4. Proposed Work

Various Bio-Medical Sensors have been deployed to track patients' vital signs, and these BMSs have been linked to the body coordinator using a STAR/MESH topology. The objective of this proposed model is to safeguard patient data like identity, illness, age, gender, and zip code from medical doctors, the general public, and paramedics. As a result, two methods for privacy preservation have been suggested in this study: K-medoid, K-anonymity, and K-means methodology, which are described below.

A. K-Anonymity Clustering Model

K-anonymity is a method for anonymization that is related to both generalization and randomization. K-anonymity is concerned with pseudo-identifier attributes. Cell suppression is used in the suggested K-anonymity method of local recoding for privacy preservation. This approach divides data into various regions and uses a mapping approach to match the categorized data to the same group's area. The advantage of using local recoding is that it prevents third-party consumers like doctors and nurses from manipulating real-time patient data by inserting incorrect values.

The suggested local recoding algorithm 2 is introduced. First, this algorithm computes each class's entire generalized domain (GD). The second move then instructs you to recalculate all of GD's potential attributes. After that, it's verified if GD is x, k anonymous by generating all generalized domains. If this is the case, the generalization domain will be used. However, before a sufficient selection is made, this procedure will be repeated and checked for all GD. The GD table based on distortion is selected if, carrying out the processes again, and the quasi-identifier (Q) scale approaches threshold levels. If not, it will proceed to the following phase, when all available qualities of GD will be measured. Ultimately, x, k-anonymity is obtained by applying the gathered generalisation domain data to a list of values. The method that implies this is known as "k-anonymity without the prior significance of the threshold k" [35].

***K-Anonymity* Algorithm:**

Input: $\boldsymbol{Q_i}$ test table;

Output: Anonymized Table;

Procedure:

Begin

N = Original Table;

$\boldsymbol{Q_i}$ = Quasi-identifier reminder of table Q = ∅;

$\mathbf{A}\boldsymbol{Q_i}$ = Anonymized Quasi-identifier table = ∅;

i=2

While $\boldsymbol{Q_i}$ is not empty {

if (find==1) {

Build a table named "$\boldsymbol{Q_i}$ bucket1" from the original table, with all variations of the chosen attributes equal in all of the table's rows;

Make a table named $\boldsymbol{Q_i}$ that contains lines that aren't in "$\boldsymbol{Q_i}$ bucket1";

}

else if (find! =1) {

Build a table from QIRT named "$\boldsymbol{Q_i}$ bucketi" with all configurations of the variables equal in all of the table's rows.

Make a table named QIRT that contains lines that aren't in "$\boldsymbol{Q_i}$ bucketi";

}

i++;

}

Return the anonymized table AT containing all the generated $\boldsymbol{Q_i}$ buckets;

End while

End

Since several studies show that k-anonymity works mainly with quasi identifier attributes, let extend our proposed approach to a test table that only includes such attributes. To begin, build a table named "bucket" in which it places identical rows based on the attributes have selected. Then, except for the rows already present in the first table, can build a new table named $\boldsymbol{Q_i}$ in which it places the remaining rows from the original test table. Afterward, instead of using the initial test table may reapply the algorithm on $\boldsymbol{Q_i}$ until it is empty. And then return all of the buckets that have been produced.

The processing of both numerical and categorical quasi-identifier attributes is what distinguishes our proposed algorithm from others. Furthermore, this algorithm is unaffected by the number of rows in the data set, allowing us to use it on large data sets. It demonstrates the multiple stages of the k-anonymity method using a sample to help you identify the algorithm.

B. K-medoid Clustering Model

The cluster technique was used to determine the most efficient way of allocating services under such restrictions as people, physicians, and medical personnel. Massive regions are divided into n spaces by the cluster, which then allocates private spaces based on policy. One advantage of the cluster system is that it manages patient data effectively through access controls, minimizing information loss and preserving privacy. This paper proposes a clustering approach focused on the classification of health monitoring and data routing, main transceiver, and cluster-based restriction to access medical data. The report is sent to the body supervisor by the health analysis and data forwarding scheme, which maintains track of the health of m patients. Moreover, the structure's integrator transmits tracked data to find scores, which retransmit to the base station. The anonymity strategy's cluster-based limitation on access to medical information is the most prevalent method. K-medoid is a machine learning-based anonymity system [2].

The K-medoid machine learning approach works by identifying the closest objects in all of the data elements and assigning them to the same cluster of the data model. The anonymity system selects classes to the designated objects of the same cluster of neighbors. The anonymity algorithm is built on K-medoid, a cluster strategy that enables access to patient data based on category and identification of individuals. Many of the procedures are completed in a limited amount of time as a result of this deployment. K-medoid is a partitioning method, which divides a dataset's n data points into K nonoverlapping distinct groups called clusters, with each data point assigned to a single cluster. As opposed to the K-means clustering technique, it is

more stable and less susceptible to noise and outliers. Rather than K means average points, medoids are used as cluster centers. Furthermore, it is clear and corresponds to a limited number of steps. K medoid distinguishes each cluster by a single data point known as the cluster's medoid. Partitioning around medoid is another name for it.

A point within a cluster that has a higher average similarity to other data points in the cluster is referred to as a medoid. The aim of the K-medoid algorithm, as seen in Equation 1, is to minimize the study of dissimilarities between cluster medoid and data points in a cluster. K-medoid is classified as follows:

$$\sum_{X_n \in Y_n} \sum_{X_n \in Y_n} |X_n - Y_n| \quad (1)$$

C. K-means Clustering Model

The classic k-means algorithm will be introduced in this section, followed by the homomorphic encryption algorithm. K-means [9] is a continuous iteration method that may group participants with identical characteristics. Assume there are 'm' participants $\{x_1, x_2, \dots, x_m\}$ that need to be divided into 'h' clusters. W = W_1, W_2, W_3,..., W_h. A 'y' dimensional function vector is used to describe each participant. The following is a summary of the k-means process:

i. Initialize w = w_1, w_2, w_3,..., w_h at random, where cj is the center of cluster W_a.

ii. According to Equation (2), assign each participant x_b to the cluster core w' that is closest to them. To use the Euclidean distance as our default distance estimate in this paper.

$$w' = \text{argmin}_{w_a \in w} \|x_a - w_a\| \quad (2)$$

iii. Using Equation (3), recalculate the new cluster core w_a, where w_a is the mean of participants in W_a.

$$w_a = \frac{1}{w_a} \sum_{x_b \in w_a} x_b \quad (3)$$

iv. Steps (2) and (3) can be repeated until there is no shift or the stop conditions are achieved.

Homomorphic encryption [36] is a technique for delegating data processing without revealing the information. Paillier cryptosystem and RSA cryptosystem [37] are two encryption techniques. The Paillier cryptosystem, for example, enables a third-party cloud network to execute additional operations on encrypted files. The properties of the Paillier cryptosystem are described by Equations. (4) and (5).

$$H(x_1 + x_2) = H(x_1) * H(x_2) \quad (4)$$

$$H(i + x) = H(x)^i \quad (5)$$

The encryption function E() is used, and s is a constant.

To encrypt private data in this article, will use the Paillier cryptosystem. The following diagrams represent the encryption and decryption of the Paillier cryptosystem.

Encryption: An individual produces two keys: a public key ek(j, c) and a private key ep =α. The message x∈J can be encoded as follows:

$$H(x) = c^x w^j \bmod j^2 \quad (6)$$

Where w is a random number that achieves the condition w ∈ j^*.

Decryption: Using cipher text H(x) and the private key, the plaintext could be measured:

$$U(H(x)) = \frac{Q(H(x)^\alpha \bmod j^2)}{Q(c^\alpha \bmod j^2)} \bmod j \quad (7)$$

Where, Q (p) = $\frac{p-1}{j}$. The following equations could be obtained from the features of the Paillier cryptosystem:

$$U(H(x_1 + x_2)) = U(H(x_1) * H(x_2)) = x_1 + x_2 \quad (8)$$

$$U(H(i * x_3)) = U(H(x_3)^i = i * x_3 \quad (9)$$

5. Performance Evaluation

The simulations are run on a computer with an "Intel® Pentium® 2.80 GHz CPU and 4 GB RAM and Python". To demonstrate the feasibility of our approach, we compare the accuracy of k-medoid, k-anonymity, and the classic k-means algorithm using patient's common datasets. The dataset is for the recovery of women who had breast cancer surgery [36]. This data was gathered from 306 patient cases. The clustering findings will help a doctor treat breast cancer. To test the success of the proposed system may use accuracy, recall rates, and the F1-Measure in this section. To demonstrate how to measure the performance predictor, Haberman's survival data set classification is used as an instance.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (11)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad (12)$$

It is not systematic to test the model solely based on accuracy and recall during the evaluation period. As a result, to integrate the precision and recall rate to produce the F1-Measure score which serves as another evaluation metric. It's shown by:

$$F_1\text{- Measure} = \frac{2*Precision*Recall}{Precision+Recall} \quad (13)$$

In our analysis, the clustering findings are tested using the three criteria mentioned above. The greater the F1-Measure, the more comparable our clustering findings are to those obtained using the traditional k-means approach. It will evaluate the correctness and usability of the clustering results by presenting the performance metrics values across different datasets in the next section. Table 1 displays the results of k-medoid, k-anonymity, and k-means clustering calculated from the unencrypted private data, in comparison to the performance metrics results achieved by our technique. Performance evaluation graphs are illustrated in figure 4.

Table 1. Comparison of Performance metrics to the traditional Clustering techniques

Clustering Techniques	Precision (%)	Recall (%)	F1-Measure (%)
K-Means	99.13	98.54	98.07
K-Anonymity	97.82	96.81	96.53
K-Medoid	95.29	96.37	95.48

Figure 1 depicts the output of the clustering algorithm in terms of K values versus time. Changing the value of K causes time to vary. The output will be altered with each value. At K-medoid, the clustering algorithm will take 1.97 seconds, and at K-anonymity, it will take 0.98 seconds. As a result, K and time are linked in a roundabout way. The proposed findings are based on six different experimental K values: 10, 20, 30, 40, and 50. At K-means, which takes 0.76 seconds, there is a slight difference. The statistics overview of the number of clusters vs execution time is seen in Table 2. Figure 3 shows the comparison of execution time in different clustering techniques.

Table 2. Execution Time in Various Clustering Techniques

Clustering Techniques	Execution Time (%)
K-Means	0.76
K-Anonymity	0.98
K-Medoid	1.97

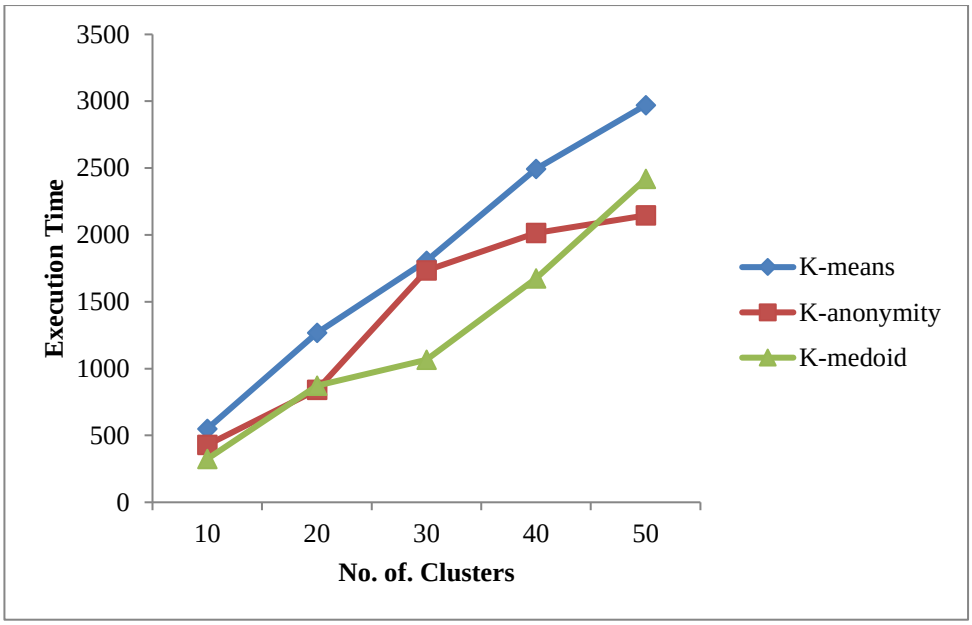

Fig. 3. Number of clusters vs execution time in various clustering techniques

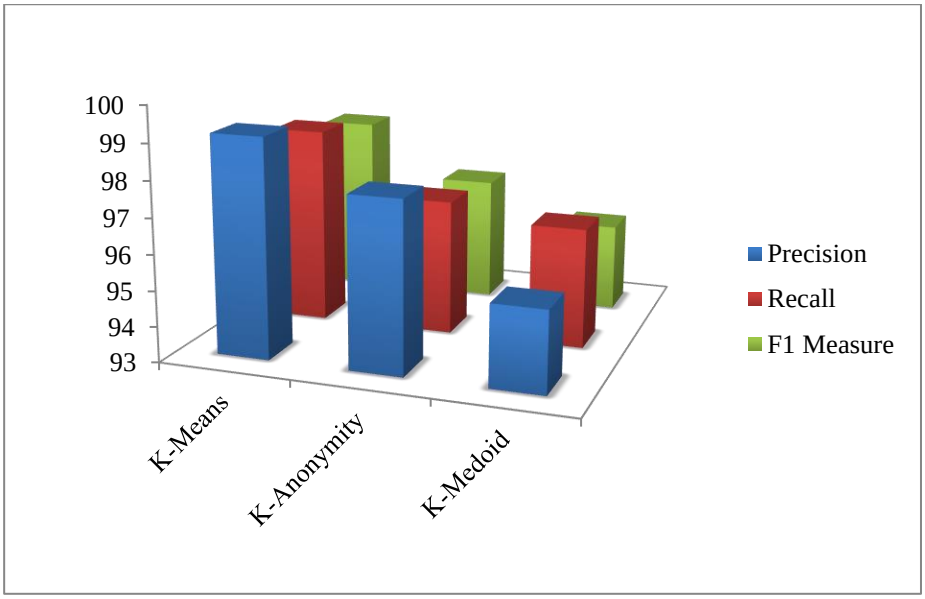

Fig. 4. Comparison of Performance Metrics

6. Conclusion

One of the most important principles in WBAN is cluster-based resource sharing using various clustering techniques. The first suggested algorithm is used to preserve personal identity by combining cluster definitions with access restriction policies. This algorithm has been successfully applied to data exchange and knowledge collection in clusters while protecting personal identity. Also, as compared to other existing clustering techniques, the highest number of clusters forming took the least amount of time. In comparison to K-medoid, K-anonymity and K mean algorithms, the next proposed algorithm has lost the least amount of knowledge. The protection of the patient's data is greatly improved as a result of these performances. In the future, the proposed algorithm for the K-means issue will be improved and compared to a supervised neural methodology for patient data protection.

References

[1] Guo, X., Lin, H., Wu, Y., & Peng, M. (2020). A new data clustering strategy for enhancing mutual privacy in healthcare IoT systems. *Future Generation Computer Systems*, *113*, 407-417.

[2] Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Security and privacy in the medical internet of things: a review. *Security and Communication Networks*, *2018*.

[3] Tang, J., Liu, A., Zhao, M., & Wang, T. (2018). An aggregate signature-based trust routing for data gathering in sensor networks. *Security and Communication Networks*, *2018*.

[4] Sun, W., Cai, Z., Liu, F., Fang, S., & Wang, G. (2017, October). A survey of data mining technology on electronic medical records. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)* (pp. 1-6). IEEE.

[5] Ullah, F., Ullah, I., Khan, A., Uddin, M. I., Alyami, H., & Alosaimi, W. (2020). Enabling Clustering for Privacy-Aware Data Dissemination Based on Medical-Healthcare-IoTs (MH-IoTs) for Wireless Body Area Network. *Journal of Healthcare Engineering*, *2020*.

[6] Xue, W., Yu, K., Hua, X., Li, Q., Qiu, W., & Zhou, B. (2018). APs' virtual positions-based reference point clustering and physical distance-based weighting for indoor Wi-Fi positioning. *IEEE Internet of Things Journal*, *5*(4), 3031-3042.

[7] Zhao, P., Jiang, H., Lui, J. C., Wang, C., Zeng, F., Xiao, F., & Li, Z. (2018). P 3-LOC: A privacy-preserving paradigm-driven framework for indoor localization. *IEEE/ACM Transactions on Networking*, *26*(6), 2856-2869.

[8] Saadatfar, H., Khosravi, S., Joloudari, J. H., Mosavi, A., & Shamshirband, S. (2020). A new K-nearest neighbors classifier for big data based on efficient data pruning. *Mathematics*, *8*(2), 286.

[9] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE transactions on pattern analysis and machine intelligence*, *24*(7), 881-892.

[10] Ortiz, A. M., Hussein, D., Park, S., Han, S. N., & Crespi, N. (2014). The cluster between internet of things and social networks: Review and research challenges. *IEEE Internet of Things Journal*, *1*(3), 206-215.

[11] Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *Ieee Access*, *5*, 8869-8879.

[12] Xing, K., Hu, C., Yu, J., Cheng, X., & Zhang, F. (2017). Mutual privacy preserving $ k $-means clustering in social participatory sensing. *IEEE Transactions on Industrial Informatics*, *13*(4), 2066-2076.

[13] Baker, S. B., Xiang, W., & Atkinson, I. (2017). Internet of things for smart healthcare: Technologies, challenges, and opportunities. *IEEE Access*, *5*, 26521-26544.

[14] Vaidya, J., & Clifton, C. (2003, August). Privacy-preserving k-means clustering over vertically partitioned data. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 206-215).

[15] Jha, S., Kruger, L., & McDaniel, P. (2005, September). Privacy preserving clustering. In *European symposium on research in computer security* (pp. 397-417). Springer, Berlin, Heidelberg.

[16] Bunn, P., & Ostrovsky, R. (2007, October). Secure two-party k-means clustering. In Proceedings of the 14th ACM conference on Computer and communications security (pp. 486-497).

[17] Xing, K., Hu, C., Yu, J., Cheng, X., & Zhang, F. (2017). Mutual privacy preserving $ k $-means clustering in social participatory sensing. *IEEE Transactions on Industrial Informatics*, *13*(4), 2066-2076.

[18] Blum, A., Dwork, C., McSherry, F., & Nissim, K. (2005, June). Practical privacy: the SuLQ framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 128-138).

[19] L.I. Yang, Guangzhou, China, research on differential privacy preserving k-means clustering, *Comput. Sci.* 1 (59) (2013) 1–34.

[20] Chen, X., Ji, J., Luo, C., Liao, W., & Li, P. (2018, December). When machine learning meets blockchain: A decentralized, privacy-preserving and secure design. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 1178-1187). IEEE.

[21] Hamza, R. (2017). A novel pseudo random sequence generator for image-cryptographic applications. *Journal of Information Security and Applications*, *35*, 119-127.

[22] Hu, H., Liu, L., & Ding, N. (2013). Pseudorandom sequence generator based on the Chen chaotic system. *Computer Physics Communications*, *184*(3), 765-768.

[23] Özkaynak, F., & Yavuz, S. (2013). Security problems for a pseudorandom sequence generator based on the Chen chaotic system. *Computer Physics Communications*, *184*(9), 2178-2181.

[24] Hua, Z., & Zhou, Y. (2016). Image encryption using 2D Logistic-adjusted-Sine map. *Information Sciences*, *339*, 237-253.

[25] Muhammad, K., Sajjad, M., & Baik, S. W. (2016). Dual-level security based cyclic18 steganographic method and its application for secure transmission of keyframes during wireless capsule endoscopy. *Journal of medical systems*, *40*(5), 114.

[26] Luo, E., Bhuiyan, M. Z. A., Wang, G., Rahman, M. A., Wu, J., & Atiquzzaman, M. (2018). Privacyprotector: Privacy-protected patient data collection in IoT-based healthcare systems. *IEEE Communications Magazine*, *56*(2), 163-168.

[27] Samani, A., Ghenniwa, H. H., & Wahaishi, A. (2015). Privacy in Internet of Things: A model and protection framework. *Procedia Computer Science*, *52*, 606-613.

[28] Sharma, S., Chen, K., & Sheth, A. (2018). Toward practical privacy-preserving analytics for IoT and cloud-based healthcare systems. *IEEE Internet Computing*, *22*(2), 42-51.

[29] Zhou, W., Jia, Y., Peng, A., Zhang, Y., & Liu, P. (2018). The effect of iot new features on security and privacy: New threats, existing solutions, and challenges yet to be solved. *IEEE Internet of Things Journal*, *6*(2), 1606-1616.

[30] Hamza, R., Yan, Z., Muhammad, K., Bellavista, P., & Titouna, F. (2020). A privacy-preserving cryptosystem for IoT E-healthcare. *Information Sciences*, *527*, 493-510.

[31] Xu, J., Xue, K., Li, S., Tian, H., Hong, J., Hong, P., & Yu, N. (2019). Healthchain: A blockchain-based privacy preserving scheme for large-scale health data. *IEEE Internet of Things Journal*, *6*(5), 8770-8781.

[32] Ullah, F., Abdullah, A. H., Kaiwartya, O., Lloret, J., & Arshad, M. M. (2017). EETP-MAC: energy efficient traffic prioritization for medium access control in wireless body area networks. *Telecommunication Systems*, 1-23.

[33] Ullah, F., Abdullah, A. H., Kaiwartya, O., & Cao, Y. (2017). TraPy-MAC: Traffic priority aware medium access control protocol for wireless body area network. *Journal of medical systems*, *41*(6), 93.

[34] Huang, M., Liu, A., Wang, T., & Huang, C. (2018). Green data gathering under delay differentiated services constraint for internet of things. *Wireless Communications and Mobile Computing*, *2018*.

[35] El Ouazzani, Z., & El Bakkali, H. (2018). A new technique ensuring privacy in big data: K-anonymity without prior value of the threshold k. *Procedia Computer Science*, *127*, 52-59.

[36] Fontaine, C., & Galand, F. (2007). A survey of homomorphic encryption for nonspecialists. *EURASIP Journal on Information Security*, *2007*, 1-10.

[37] Laurichesse, D., & Blain, L. (1991). Optimized implementation of RSA cryptosystem. *Computers & Security*, *10*(3), 263-267.

[38] T.-S. Lim, Haberman's survival data set, 1999, http://archive.ics.uci.edu/ml/datasets/Haberman.