The Journal of Supercomputing
https://doi.org/10.1007/s11227-021-04173-w

Privacy preserving framework using Gaussian mutation based firebug optimization in cloud computing

K. Anand[1] · A. Vijayaraj[2] · M. Vijay Anand[3]

Accepted: 25 October 2021

Abstract

In recent years, the data exchange among the service providers and users has been increased tremendously. Various organizations like banking sectors, health as well as government associations collect and process the data regarding an individual for their beneficial purpose. However, data confidentiality and data privacy are still considered as significant challenges while sharing sensitive data. The cloud storage servers based on unencrypted data are susceptible to both external and internal attacks established by strangers or untrustworthy cloud service providers. Since the medical data are sensitive, the risk based on privacy enhances at the moment of subcontracting entity medical records to the cloud. The significant intention of the proposed approach involves securing and preserving sensitive healthcare data. Here, data hiding and data restoration operations are considered as two significant operations of the proposed framework. Initially, an optimal key is generated in the data hiding operation. This paper proposes a Gaussian mutation-based firebug optimization (GM-FBO) algorithm for the generation of an optimal key. The experiments are conducted using three different healthcare datasets, namely HPD, Medical MIMIC-III, and MHEALTH. The efficiency of the proposed model is compared with different state-of-the-art techniques to determine the efficiency of the system.

Keywords Health care data · Cloud computing · Gaussian mutation · Firebug optimization · Privacy

✉ K. Anand
anandonmail@gmail.com

[1] Associate Professor, Center for Artificial Intelligence, Chennai Institute of Technology, Kundrathur, Chennai, India

[2] Associate professor, Department of Information Technology, Vignan's Foundation for Science Technology and Research, Vadlamudi, Guntur, Andhra Pradesh, India

[3] Professor, Department of Computer Science and Engineering, Saveetha Engineering College, Thandalam, Chennai, India

Published online: 17 January 2022

1 Introduction

Cloud computing is an emerging computing platform that provides immense support for various fields mainly health care sector, business and education with large scalable framework and unlimited internet access. Due to reliable and secure atmosphere of cloud, many users prefer cloud service provider platforms [1]. However, data transmission using the cloud faces various risks regarding privacy and security management. Even though cloud offers numerous benefits, a few complexities like byzantine fault, malicious attacks, data leakage, hijacking, service attacks, and other technology vulnerabilities will cause an impact on data integrity [2]. In order to know the status of data stored in cloud, the cloud auditing is essential to be established. In recent years, data encryption is performed by utilizing encryption algorithms that convert the text language into ciphertext and this type of data is not allowed to be read by an unauthorized person [3].

Similarly, an individual key is provided to the user which helps the authorized user to read the original text. Furthermore, privacy management in the cloud consists of two phases; they are storage security and processing security [4]. In this, data storage security incorporates data confidential related problems when the data gets saved in the cloud centre, while data processing security incorporates the problems related to data confidential runtime. When once the data on the cloud are exhibited by third-party users, the abuse of private data cannotbe stopped [5]. Now, the users are very aware of security intrusions involved in privacy preservation issues. In recent times, cloud computing platforms are growing rapidly in the health field for the secure storage of sensitive health records in cloud services [6]. The location-based data such as time, important sign readings are transferred to privacy providers for granting high quality services so that the patients can be saved in case of an emergency situation. Cloud providers depend on virtual resources that are functioned through the randomly distributed network structure [7]. Here, data loss and data leakage are intolerable problems, mainly in the field of health care sector. Because of more accessible information, notably in the health sector, machine learning techniques are already adopted [8].

The health care specialist, as well as patients, can obtain their health records from remote locations in required situations by using cloud services. By this, timely and reliable data sharing without duplicating data offers precise treatment to the patients [9]. Therefore, the encryption of confidential data is necessary before storing the data in the cloud service. At last, the sensitive data are authenticated and authorized in the cloud for secure data preservation. Moreover, many existing techniques based on security and privacy preservation pose various drawbacks in the e-healthcare framework [10]. Therefore, a novel method is to be introduced for secure data transmission and storage in big data analysis. The cloud service has to preserve privacy on both customer side and service side. Various existing data transferring techniques face many difficulties in the e-health care sector [34–48]. In this paper, the significant intention of the proposed approach involves in securing and preserving the sensitive healthcare data. The major contributions of the paper are delineated below.

- A novel Gaussian mutation based firebug optimization (GM-FBO) algorithm is proposed to generate optimal keys.
- The data hiding and data restoration operations are employed to perform an effective privacy preserving based healthcare design.
- The comparative analysis is carried out for the proposed model and various other techniques thereby determining the effectiveness of the system.

The remaining paper is arranged in the following manner. Section 2 depicts the few literature works based on privacy and securing management. In section 3 the problem statement for healthcare system is presented. The proposed methodology constituting data hiding and data restoration operations are discussed. In Sect. 4, the evaluation performances and the comparative analysis are performed. Finally, in Sect. 5, the conclusion and the future directions of the paper are presented.

2 Review of related works

Numerous research works are carried out regarding the privacy preservation concept in cloud computing. However, they create some disadvantages like low convergence time and speed, poor searching capability, less precision, reduced robustness, subject to malicious attacks, minimized security rate, etc. Some of the works concerning privacy and security management are discussed in Table 1.

Alphonsa et al. [11] suggested the genetically modified glow-worm (GMGW) swarm optimization algorithm-based privacy preservation model in cloud. This technique increased the accuracy level of both the data restoration as well as data sanitization. The statistical analysis was performed in terms of best, worst, mean, median and standard deviation also comparative analysis was carried out and the results revealed that the GMGW algorithm performed better than other compared methods.

Wu et al. [12] presented the edge cloud computing paradigm based private random decision tree (PRDT) framework for better privacy preservation and data utilization process. The comparative analysis was performed and illustrated that prediction accuracy was higher in PRDT-DRU (data repeated usage) method but creates complications in data utilization process.

Abdo et al. [13] developed the mobile health care monitoring system based on cloud computing along with location privacy preservation approach. The performance of the approach was evaluated in terms of reliability, scalability, efficiency, privacy and security, while the analysis result proved it managed the trade-off between data utility and location privacy. But robustness gets reduced when the system was scaled to some extent.

Khatiwada et al. [14] demonstrated the access control (AC) model for privacy preservation in health care sector. This approach was comprised of six phases and each phase carries separate tasks for secured data transmission. The performance of AC model was validated by using metrics like memory usage, detection rate and compared with existing methods. The simulation outcome showed that it achieved 91% of data rate than other methods but was non-flexible for GDPR.

Ahamad et al. [15] introduced the Jaya-shark smell optimization (J-SSO) algorithm for privacy preservation in cloud computing process. This technique solves optimization problems, obtains superior value in flexibility and exploration capability. However, this technique arises complexities like low convergence speed and high time consuming.

Mondal et al. [16] presented the improved honeypot cryptographic approach to preserve data privacy in cloud against data attacks. In this approach, extraction and classification were executed by using Grey level co-occurrence matrix and convolutional neural network classifier, respectively. The performance analysis was carried out in terms of accuracy, recall and f1 score. The outcome of this technique proved that the security rate was improved with 95.3% accuracy.

Yang et al. [17] implemented the stateless cloud auditing technique intended for privacy preserving process in non-manager dynamic group data. The stateless cloud auditing technique decreases the overhead computation cost, and performance was measured in terms of functionality, security and cost. This technique efficiently secures the data and also lowers the computation cost while there was no consideration of batch auditing.

Fang et al. [18] demonstrated the efficient machine learning based federated learning (FL) approach for privacy preservation in the cloud server. This approach enhances the training efficiency, lessens execution time and reduces the communication cost. The simulation measures were performed and the result revealed that the FL approach achieved greater accuracy but created inference on output.

Jayaram et al. [19] suggested the secure edge cloud-based healthcare system (SECHS) for efficient prediction and privacy preservation process. Also, disease detection and rehabilitation approach in healthcare were performed by utilizing adaptive weighted probabilistic classifier. The SECHS approach was compared with existing technique to validate the performance and showed that SECHS obtained greater predicting time and accuracy but failed to track an edge-to-edge object.

Prabha et al. [20] introduced the suppressed K-anonymity multi-factor authentication-based Schmidt cryptography (SKMA-SC) method for handling the data in a secured state in cloud storage. This technique was performed by three steps, namely data registering, authenticating and accessing processes and the evaluation result showed that it had enhanced privacy and decreased computation complexity but reduced data integrity.

3 Problem statement

Even though, the cloud technology comprises immense benefits, data confidentiality and data privacy are still considered as a significant challenge during the utilization of cloud storage services. The cloud storage servers based on unencrypted data are susceptible to both external and internal attacks established by strangers or untrustworthy or dubious cloud service providers. Since the medical data are sensitive, the risk based on privacy enhances at the moment of subcontracting entity medical records to the cloud. Even though the cloud computing offers secured services it is prone to external threats. Hence, modelling a privacy-preserving and secure health data is considered as a major challenging problem and it needs to be solved. In order to provide better and fine healthcare services, quick access of healthcare data is necessary [33]. This further provides high life quality and on-time treatment during emergencies. In recent years, the total number of hospital occupancies is considerably minimized due to the establishment of e-healthcare systems. Therefore, medical purpose sensitive data are involved mostly in the cloud service system. Furthermore, cloud computing technology constitutes certain benefits like better content storage and application hosting as well as minimum consumption expenses. But effective data extraction and analysis are considered as a challenging task. To overcome the above-mentioned drawbacks, a Gaussian mutation-based firebug optimization (GM-FBO) algorithm for preserving privacy in the cloud service system regarding healthcare is proposed in this paper.

4 Proposed approach

The architectural diagram for the proposed approach is shown in Fig. 1. The significant intention of the proposed approach involves securing and preserving sensitive healthcare data. Data hiding operation and data restoration operation are considered as two significant operations of the proposed framework. Initially, an optimal key is generated in data hiding operation. In this paper, an optimal key is generated using a Gaussian mutation-based firebug optimization (GM-FBO) algorithm. The data which are to be preserved are said to be known as secure data are then sent to the receiver. Once the secure data are received by the receiver, the original data can be viewed only if a similar key is provided to the receiver. Therefore, the original hidden sensitive data are recovered by the receiver by using an inverse optimal key which comes under restoration operation. The original data received are then utilized for analysing and diagnosing medical reports, and the data are then sent to the particular person (can be caretaker, patient or other authorized persons) in case of emergency. The step-by-step process involved in the proposed technique is discussed in the upcoming sections.

4.1 Fitness function evaluation

The significant intention of the proposed approach involves securing and preserving the sensitive healthcare data, and the main objective of this paper is to accomplish an optimal key and it can be done by employing a hybrid algorithm named GM-FBO algorithm. Here, the key is considered as an input solution to the GM-FBO algorithm. The mathematical formulation involved in deriving the fitness function is stated as follows:

$$\text{Fitness function} f = \text{Minimum}\,(\text{Key size } (\delta_k)) \tag{1}$$

From Eq. (1),

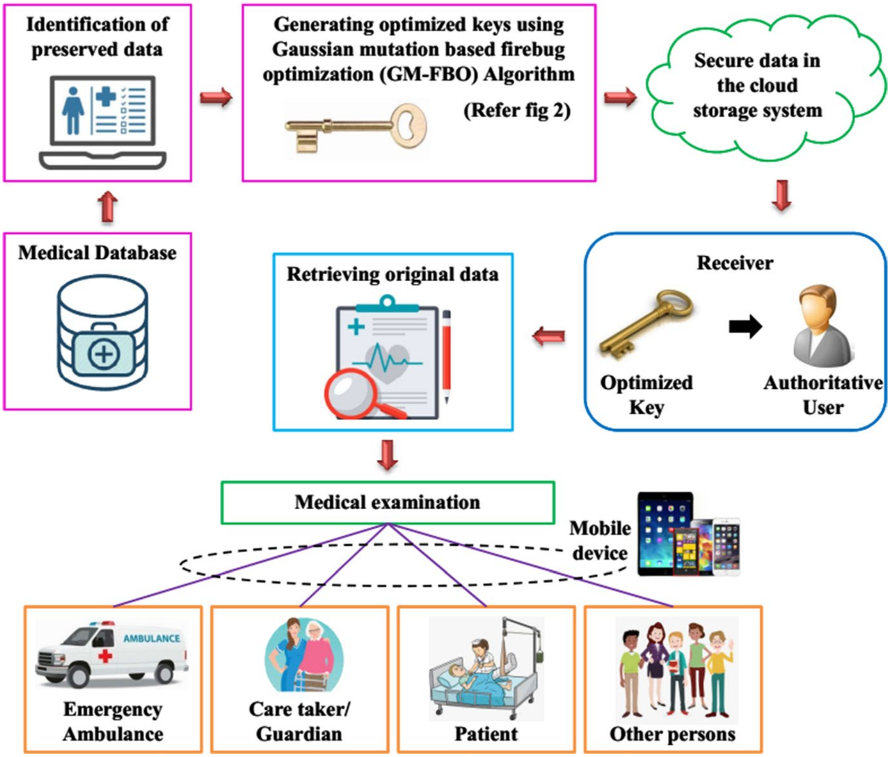


Fig. 1 Proposed Workflow

$$\delta_k = \frac{\sum_{k=1}^{n} S_D}{\sum_{k=1}^{n} O_D} - \left[\frac{\sum_{k=1}^{n} O_D - \sum_{k=1}^{n} P_D}{\sum_{k=1}^{n} O_D} \right] \tag{2}$$

From Eq. (2), the original data are denoted by O_D. n signifies the total number of medical data. The preserved data and the secure data are represented by P_D and O_D, respectively [15].

4.2 Data hiding operation

The data hiding operation in other words referred to as a data preservation technique in which the sensitive healthcare data are preserved by employing an optimal key [21]. In this paper, an optimal key is generated using the GM-FBO algorithm that is depicted clearly in the following subsection. Initially, an optimal key is changed into binary values so as to hide the data. The healthcare data are then proliferated from the obtained binary values referred to as secure data. The steps involved in obtaining binary data from an optimal key are discussed below. Let us assume, $\delta_1 \times \delta_2$ be the data size; where δ_1 are the records and δ_2 is the fields. The secure data are generated by multiplying the optimal key size of about 20×1 with the original data. The length of the binary values and the original data is analogous during the conversion process. The elements present in 20×1 are categorized into five batches and each constitutes four elements and 40 binary bits are obtained by converting every element thereby forming five 40×4 data. Thus, the obtained five 40×4 data are linked to obtain a total aggregate with the size of about 200×4. Soon after obtaining the binary data, secure data are obtained by multiplying ith the original data.

4.2.1 Gaussian mutation-based firebug optimization (GM-FBO) for optimal key generation

In this section, a GM-FBO algorithm is employed in generating an optimal key. The Firebug swarm optimization algorithm along with Gaussian mutation operation and cross over operation of a genetic algorithm is proposed to conquer the drawbacks like poor convergence rate and minimum global search capability of the original FBO algorithm. In addition to this, the selection capability of the FBO algorithm is enhanced by using mutation and cross over operation. A detailed description of the GM-FBO algorithm is discussed in the upcoming section.

4.2.1.1 Biological characteristics The biological name of firebug is Pyrrhocoris apterus which belongs to an insect family. These firebugs constitute two significant behaviours; one is the roaming behaviour and the other is the exploring behaviour. The motion of firebugs that attempts to determine the best mating partner can be viewed naturally as an optimization approach. Since the characteristic behaviour of firebugs is complex; studying the behaviour is considered as relevant and interesting research area. The following sub-section depicts a few interesting behaviours of firebugs [22].

- The male firebug attracts and defends the colonies of the female firebugs using chemical signals by means of pheromones.
- Only the fit female bugs are selected by the male firebug for the process of mating.
- Both female and male firebugs that mates successfully form a tandem and be with each other over a particular time duration.
- The male firebugs defend the female colonies with the highest fitness value are attracted by the female firebugs.
- Each firebug does not disperse from the group; meanwhile, they move together as an aggregation.

The behaviours and the respective mathematical expression involved in the GM-FBO algorithm are delineated as follows.

4.2.1.2 Forming female colonies As mentioned above, only the fit female bugs are selected by the male firebug and the significant objective involves minimizing the cost function since fittest firebugs are associated with minimum cost. Here,n_f represents the number of female bugs which is distributed randomly in the search area and n_m denotes the number of male bugs which is distributed randomly in the search area. Each bug constitutes a real scalar cost with respect to the fitness value and a position value. Otherwise stated, the initial female firebug position is regarded as a uniform random vector variable in the search space area [23].

4.2.1.3 Selection of mates The matting process takes place between the male firebug and the fittest female firebug in the colony. The location of every male bug is initialized to the fittest female firebug's location. The significant differences among the firebug optimization and other optimization algorithms are that the FBO utilizes an element wise operation referred to as Hadamard matrix multiplication for updating the position. The females in a colony are parallelly updated. In this paper, the selection capability of FBO algorithm is enhanced by using Gaussian mutation operation and cross over operation which is discussed below.

- *Crossover operation* The most significant phases in the genetic algorithm are the crossover operation. The cross over operation permits the integration of genetic factors containing one or more solutions. Certainly, majority of the groups comprises two parents. This operator implements purposes where the parents with respect to genetic facts are associated with it [24].
- *Gaussian mutation operation* The Gaussian mutation operation [25] in other words stated as random vector addition that obeys the Gaussian distribution function or normal distribution function. This Gaussian mutation operation plays a vital role in probability theories and mathematical statistics. If a random integer y complies Gaussian distribution by numerical expected value η and variance σ^2 then they are represented as $P(\eta, \sigma^2)$.

$$F(y) = \frac{1}{\sqrt{2\pi\sigma}} Exp\left(\frac{-(y-\eta)^2}{2\sigma^2}\right) \tag{3}$$

From Eq. (3), $F(y)$ signifies the probability density function, η indicates the expected value which identifies the position and σ depicts the standard deviation which identifies the magnitude. Numerous random features and phenomenon in the natural world are approximately described with normal distribution function.

4.2.1.4 Female firebugs based on chemotactic motion After initializing, the female bug location is updated. Every female firebug moves to the dominant firebug to attain effective matrix operation. Consider $(r(s).f$ as B with n_f matrix and its columns correspond to positions of female bugs. Employing the Hadamard multiplication operations, all the female bugs are grouped in a particular colony and their respective numerical notations are given as follows,

$$Q_x \leftarrow Rm(r(s).x, 1, n_f) \tag{4}$$

$$Q_y \leftarrow Rm(r(t).x, 1, n_f) \tag{5}$$

From the above equations, t implies a random number that ranges between $[1, n_f]$, $Rm(T, s, w)$ indicates the return matrix includes s copies of T in row matrix and w copies of T in column matrix dimensions. Hence, if the matrix is $l \times m$ dimension, then $Rm(T, s, w)$ returns $sl \times wm$ matrix.

$$r(s).f \leftarrow r(s).f + c^1\Theta(Q_x - r(s).f) + c^2\Theta(Q_y - r(s).f) \tag{6}$$

In Eq. (6), c^1, c^2 denotes the matrix cost function.

4.2.1.5 Male bugs attracting to the female bugs The male bug attracts towards the fit female bugs and not their colonies. Avoiding the competitiveness among diverse male firebugs permits both the male and female colonies linked with the male firebugs thereby exploring a wide search space area without being inspired to the similar best location of the fit female firebugs.

$$r(s).x \leftarrow r(s).x + c^3\Theta(q - r(s).t) \tag{7}$$

4.2.1.6 Swarm organization In swarm cohesion process, each firebug does not disperse from the group; meanwhile, they move together as an aggregation. The swarm takes stochastically motion so that a single bug cannot move in the direction opposite to the direction of the swarm. The mathematical expression based on random motion of swarm is depicted in Eq. (8)

$$r(s).x \leftarrow r(s).x + c^4\Theta(q - r(a).t) \tag{8}$$

The male bugs move towards the fitness female bugs and are illustrated using the below equation,

$$r(s).x \leftarrow r(s).x + \beta(q - r(s).x) \tag{9}$$

In a similar way, the movement of strong female bugs towards alpha males and weak female bugs towards subordinate males are carried out and are mathematically expressed in the below equation,

$$r(s).f \leftarrow r(s).f + C^1\Theta(Q_x - r(s).f) + C^2\Theta(Q_y - r(s).f) \tag{10}$$

From the above equation, $C^1\Theta(Q_x - r(s).f)$ and $C^2\Theta(Q_y - r(s).f)$ indicates the attraction of female bugs towards dominant males and subordinate male bugs, respectively, and the attraction strength was found using C^1 and C^2 matrixes. The flow chart representation for the GM-FBO technique is illustrated in Fig. 2.

4.3 Data restoration operation

The data restoration scheme constitutes two different types of data, namely the index data and the sensitive data [26]. At first, the vector form of secure data with compatible length is generated by considering the sensitive data. Then the key index and the sensitive data are multiplied, and the value is then added with the secure data

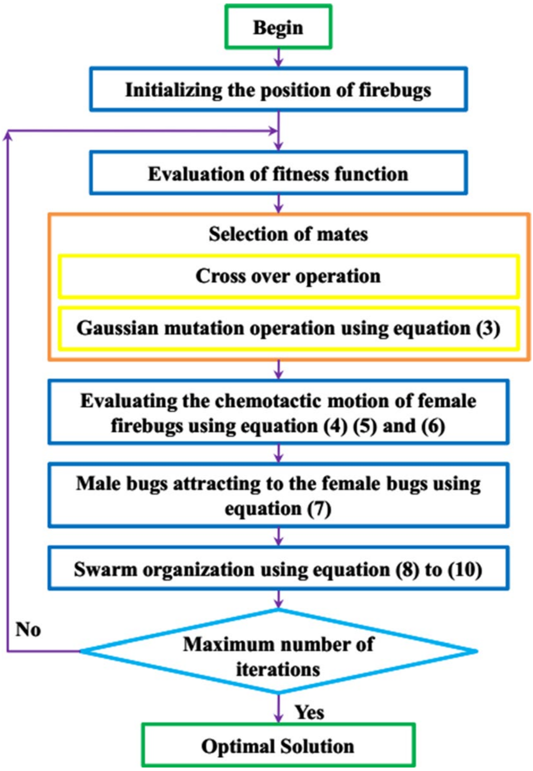

Begin
Initializing the position of firebugs
Evaluation of fitness function
Selection of mates
Cross over operation
Gaussian mutation operation using equation (3)
Evaluating the chemotactic motion of female firebugs using equation (4) (5) and (6)
Male bugs attracting to the female bugs using equation (7)
Swarm organization using equation (8) to (10)
No
Maximum number of iterations
Yes
Optimal Solution

Fig. 2 Flow chart representation of the proposed GM-FBO approach

for acquiring the original data or restored data. Let us assume the sensitive data as 4. Firstly, the vector and the generated optimal key is multiplied which are generated for the sensitive data (4). Finally, the obtained result and the secure data are added to obtain the original data. The successful restoration of original data takes place only if the optimal key generated using GM-FBO is accurate. On the other hand, if the generated key is not optimal, then the restoration process of the original data is unable to accomplish effectively. Hence, the correlation coefficient of recovered and the original data are resolved thereby demonstrating the efficiency of the proposed GM-FBO technique [26].

5 Experimental results and analysis

The experiments are simulated using the Matlab 2020b software and the cloud environment is simulated using a CloudSim3.0.3 tool. The experiments are conducted in a PC equipped with an 11th Generation Intel Core i5-1135G7 Processor, Windows 10 Home 64 OS, and 16 GB Soldered DDR4 3200 MHz memory. Three healthcare datasets were taken to conduct the experiments, namely: Healthcare problem dataset (HPD) [27], Medical Information Mart for Intensive Care (MIMIC-III) [28], and

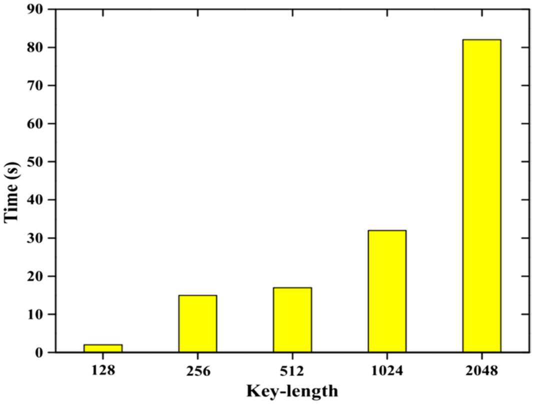
Mobile HEALTH (MHEALTH) [29]. A brief description of these datasets is presented as follows:

- *HPD* This dataset is mainly used to predict the stroke symptoms in heart disease patients. The total number of attributes present in the dataset is 12. The patient's id, gender, age, presence of hypertension, presence of heart disease, married, type of job, type of residence, average glucose level, body mass index, and stroke history are the attributes present in the dataset. The attributes that are secured are, namely age, hypertension, type of job, BMI, average glucose level, and smoking status.
- *MIMIC-II* It is a large single-centre database that consists of the information of the patients admitted into the critical care unit. The information includes medications, laboratory measurements, vital signs, notes, and observation reports, imaging reports, length of stay, diagnostic codes, etc. The data that need to be secured in this dataset are, namely noted and observation report, laboratory measurements, imaging reports, vital signs, and diagnostic codes.
- *MHEALTH* This dataset consists of the physical activity information obtained from ten volunteers. The different motions exhibited by the various parts of the body are collected via different sensors. There are 12 activities and 24 attributes present in the dataset. The 24 attributes present in the dataset are secured using the proposed methodology.

5.1 Experimental results

The overall time taken by the proposed GM-FBO algorithm by varying the size of the secret key and different patient records is presented in Fig. 3. The key generation process was repeated for a total of 10,000 patients records which are equally distributed among five virtual machines. The time is taken by the proposed methodology for different virtual machines in the range 3–18 and a secret key size of 1024 bits is presented in Fig. 4. Figure 5 presents the key size variation for different instances present in dataset-I.

Fig. 3 Time taken for key generation by varying the key length



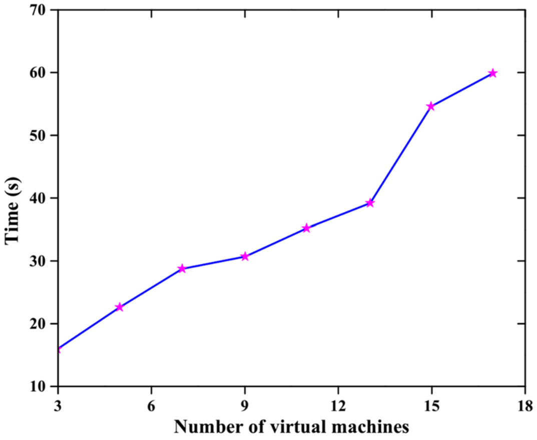


Fig. 4 Computation of key generation time by varying the number of virtual machines

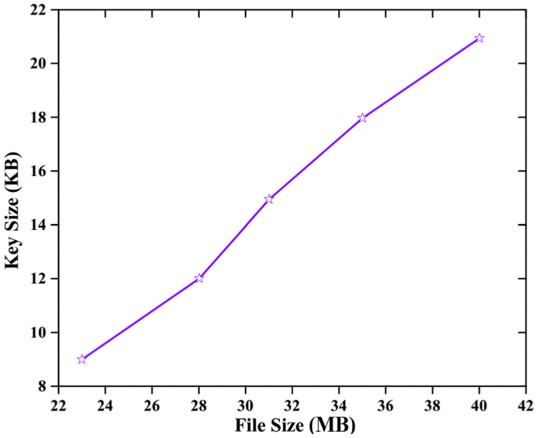
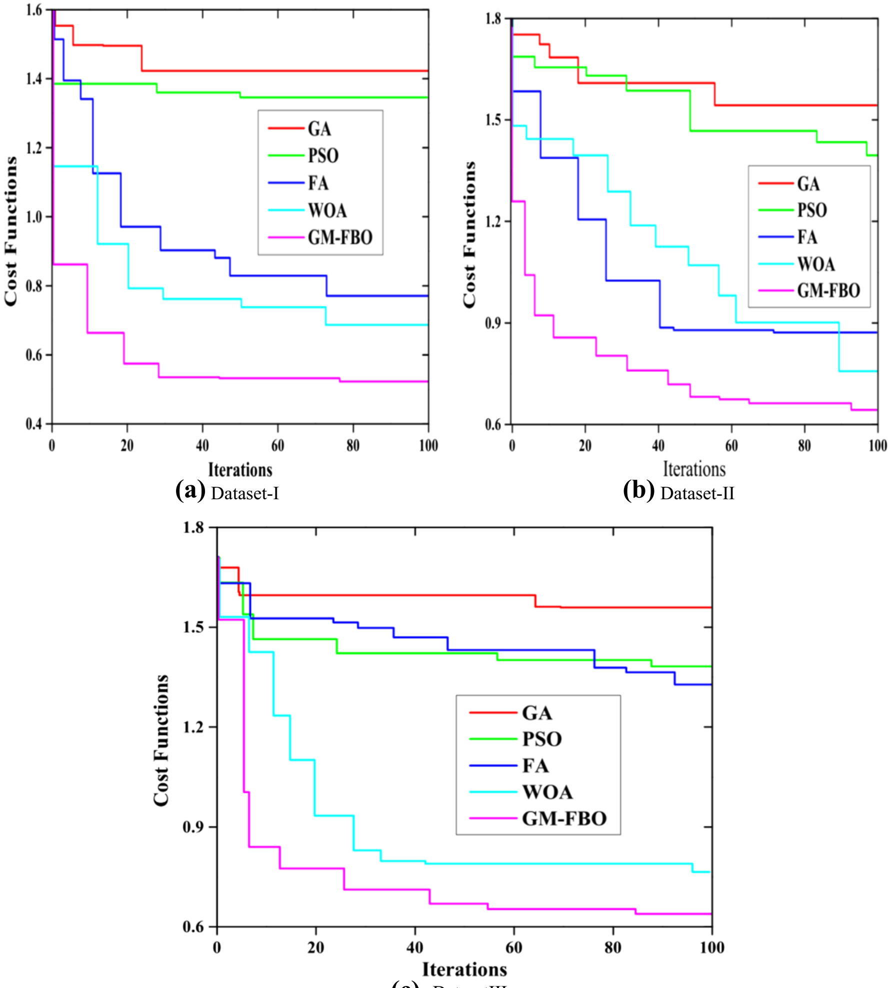


Fig. 5 Random key size variation to increase the privacy

The performance of the proposed GM-FBO algorithm is evaluated using different measures, namely privacy and fitness. The results obtained by comparing with the FBA algorithm are presented in Table 2. The GM-FBO algorithms key generation process improves the fitness value. The key sizes are varied from 100 to 500 for every measure. The privacy measure mainly denotes the percentage of privacy improved by the proposed methodology for the cloud computing environment. The main aim of the privacy measure is to improve the privacy offered to the user's healthcare information. The proposed algorithm offers a privacy measure of 0.49, 0.50, 0.52, 0.40, and 0.42 for key size of 100, 200, 300, 400, and 500, respectively. The privacy measure is relatively high when compared to the FBA algorithm. The fitness value is mainly based on the objective function (optimal key with minimum length) that needs to be minimized. The FBA algorithm offers a fitness measure of 0.35, 0.50, 0.48, 0.42, and 0.40 which is relatively higher than the proposed GM-FBO algorithm. Hence it is proven that the proposed methodology offers low fitness which is effective when generating an optimal key with the minimum length for different medical datasets.

5.2 Performance evaluation using convergence

The optimal key obtained using the proposed GM-FBO algorithm is the best key obtained for the secure generation process. The convergence performance is said to be higher when the cost reduces as the number of iterations increases. The comparative analysis in terms of comparing the proposed method to the existing system is present in Fig. 6. The model is run for a total of 100 iterations. For the comparative analysis, a total of three datasets is taken and the methods taken for comparison



(a) Dataset-I

(b) Dataset-II

(c) DatasetIII

Fig. 6 Convergence analysis of the proposed methodology with existing techniques

Table 1 Outline of related works

Author name/year	Adopted techniques	Databases	Advantages	Challenges
Alphonsa e tal. [11]	GMGW algorithm	Heart disease dataset	Preserve privacy of data	Less reliability
Wu et.al [12]	PRDT-DRU method	Adult and nursery dataset	High prediction accuracy,	High computational overhead
Abdo et al. [13]	Mobile health care monitoring system	Health dataset from online sources	Manages tradeoff between data utility and location privacy	Less robustness
Khatiwada et al. [14]	AC model	Heart disease dataset	Control the unauthorized access of data	Inflexible, high cost
Ahamad et al. [15]	J-SSO algorithm	UCI repository datasets	Highly flexible, enhances exploration capability	High time consuming, low convergence speed
Mondal et al. [16]	Improved honeypot cryptographic approach	KDD Cup 99 dataset	Higher security rate	High computational overhead
Yang et al. [17]	stateless cloud auditing technique	Decentralized dataset	High data security, low computation cost	No consideration of batch auditing
Fang et al. [18]	FL approach	UCI Human Activity Recognition Dataset	Enhances the training efficiency, lessens execution time, reduces the communication cost	Creates inference on output
Jayaram et al. [19]	SECHS method	Parkinson disease dataset	High prediction time, high prediction accuracy	Failed to track edge to edge object
Prabha et al. [20]	SKMA-SC method	Amazon Access sample Dataset	Provide enhanced data restoration	Reduces data integrity

Table 2 Comparison using privacy and fitness measures

Key size	Privacy measure		Fitness measure	
	FBA	Proposed GM-FBO	FBA	Proposed GM-FBO
100	0.10	0.49	0.35	0.27
200	0.39	0.51	0.50	0.28
300	0.39	0.52	0.48	0.31
400	0.18	0.40	0.42	0.29
500	0.20	0.42	0.40	0.32

are Genetic Algorithm (GA) [24], Particle Swarm Optimization (PSO) [22], Firefly algorithm (FA) [23], and Whale optimization algorithm (WOA) [24]. For the three datasets, the proposed algorithm has a minimal cost function and it also converges with a minimal cost function at the end of the 100th iteration. The GA and PSO algorithm have the highest cost during convergence when compared to the remaining algorithms. The FA and WOA algorithm's performance degrade with increasing search space. The proposed GM-FBO algorithm offers optimal outcomes for the three datasets. Hence, we can conclude that the proposed model is efficient in reducing the cost function when compared with the existing techniques.

5.3 Chosen plaintext and known plaintext attack analysis

The proposed methodology is evaluated in terms of known plaintext attacks (KPA) and Cipher plaintext attacks (CPA). The KPA is a cryptanalysis attack model where the attacker gains access to both the plaintext and ciphertext. The main aim of this attacker is to disclose the secret key used. The KPA attack is conducted by either identifying the statistical relationship that exists between a single data instance to the overall data or either identifying the relationship of a single secret data instance with the overall secret data. The chosen plaintext attack (CPA) is also a cryptanalysis attack model used by the attacker to gain access to the ciphertexts generated by random plaintexts. The main aim of the attacker is to minimize the security of the encryption scheme used. The CPA attack is conducted by analysing the statistical significance between the secret data and its appropriate plain text that is restored.

The results obtained for both the KPA and CPA attacks are shown in Tables 3 and 4. In KPA attack analysis, the proposed methodology offers improved performance of 8.5%, 5.9%, 6.9%, 4.5%, 2.9%, and 1.9% when compared to the state-of-the-art techniques such as SKMA-SC [10], PRDT-DRU [2], J-SSO [5], Improved honeypot [6], and SECHS [9]. Based on the performance shown in Table 4, we can observe that the proposed GM-FBO is very reactive against the CPA attacks when compared to the state-of-the-art techniques and the main reason is the integration of the GM mechanism with the FBO algorithm. Even though the state-of-the-art techniques offer optimal performance, they suffer from different complexities such as high computational time, complexity in handling a large number of data instances.

Table 3 Comparison using KPA attacks

Techniques	Dataset-I	Dataset-II	Dataset-III
SKMA-SC [10]	0.9024	0.9061	0.9014
PRDT-DRU [2]	0.9259	0.9291	0.9214
J-SSO [5]	0.9158	0.9195	0.9154
Improved honeypot [6]	0.9395	0.9395	0.9314
SECHS [9]	0.9559	0.9595	0.9547
GMGW [1]	0.9654	0.9654	0.9654
Proposed GM-FBO	0.9851	0.9974	0.9956

Table 4 Comparison using CPA attacks

Techniques	Dataset-I	Dataset-II	Dataset-III
SKMA-SC [10]	0.9256	0.9364	0.9347
PRDT-DRU [2]	0.9365	0.9324	0.9365
J-SSO [5]	0.9244	0.9255	0.9245
Improved honeypot [6]	0.9547	0.9596	0.9514
SECHS [9]	0.9485	0.9485	0.9356
GMGW [1]	0.9546	0.9484	0.9584
Proposed	0.9984	0.9984	0.9958

Table 5 Statistical Analysis results for Dataset-I

Techniques	Best	Worst	Mean	Median	Standard deviation
GA	1.425546	1.45962	1.46521	1.45684	0.019599
PSO	1.456985	1.51595	1.49568	1.45698	0.015995
FA	1.459852	1.45995	1.36548	1.41559	0.042648
WOA	0.519526	0.78955	0.66589	0.69854	0.084578
Proposed GM-FBO	0.37458	0.69588	0.55998	0.65487	0.12545

5.4 Statistical analysis

The stochastic nature of the meta-heuristic algorithm often deters their performance when achieving optimal results. The statistical analysis is mainly conducted by taking the mean, best, median, and standard deviation of different metaheuristic algorithms such as Genetic Algorithm (GA) [24], Particle Swarm Optimization (PSO) [30], Firefly algorithm (FA) [31], and Whale optimization algorithm (WOA) [32]. Every algorithm has been executed a total of 5 times. Table 5, 6, 7 shows the results obtained for the HPD (Dataset-I), MIMIC-II (Dataset-II), and MHEALTH (Dataset-III) datasets. Each table shows the analysis conducted in terms of mean, best case, worst case, median, and standard deviation.

Table 6 Statistical Analysis results for Dataset-II

Techniques	Best	Worst	Mean	Median	Standard deviation
GA	1.50458	0.80145	1.501548	1.51487	0.04858
PSO	1.52648	1.50654	1.59652	1.58694	0.02998
FA	1.32553	1.65487	1.41547	1.40154	0.05895
WOA	0.69584	1.55648	0.78959	0.70958	0.04568
Proposed GM-FBO	0.59688	1.188882	0.80648	0.69581	0.25684

Table 7 Statistical Analysis results for Dataset-II

Techniques	Best	Worst	Mean	Median	Standard deviation
GA	1.59852	1.65875	1.62548	1.62548	0.02854
PSO	1.63552	1.70154	1.6789	1.54585	0.05844
FA	1.45594	1.6015	1.5469	1.545958	0.05689
WOA	0.665858	0.80558	0.73658	0.75956	0.05954
Proposed GM-FBO	0.38547	0.94587	0.74589	0.8145	0.23258

For dataset-I, the proposed GM-FBO offers an improvement of 33.25, 71.25, 74.65, and 74.65% when compared to the GA, PSO, FA, and WOA algorithms. The superiority is also visible both in the worst-case scenario for the three datasets shown in Tables 4, 5, and 6. The mean value of the proposed model for dataset-II is 42.36, 49.36, 47.32, and 50.36% higher than the GA, PSO, FA, and WOA algorithms as shown in Table 6. Table 7 also shows superiority to the state-of-the-art techniques in terms of best case, worst case, and median values. The GA, PSO, and FA algorithm suffers from premature convergence and is often stuck in the local optima. The WOA algorithm suffers from the low convergence speed, hence its performance degrades. The genetic crossover and Gaussian mutation operation enhance the capability of the FBO algorithm with increased convergence speed.

5.5 Data restoration efficiency

The statistical relationship between the actual and restored data is analysed via the data restoration efficiency. Tables 8, 9 and 10 present the data comparison results of the data restoration efficiency when evaluated with the three state-of-the-art techniques in terms of the three datasets. The techniques taken for comparison are, namely SKMA-SC [10], PRDT-DRU [2], J-SSO [5], Improved honeypot [6], and SECHS [9]. From the results shown in Table-7, we can analyse that there is an improvement of 3.21, 6.32, 5.89, 5.78, and 6.02% over the SKMA-SC, PRDT-DRU, J-SSO, Improved honeypot, and SECHS techniques for a total of 10 instances.

Based on the results shown in Table 8, we can see an improvement of 5.43, 4.8, 4.46, 4.82, 4.55, and 4.13% of the proposed methodology over the existing techniques

Table 8 Comparison of the data restoration effectiveness with the state-of-the-art techniques using dataset-I

Techniques	Number of instances taken from the input dataset									
	1	2	3	4	5	6	7	8	9	10
SKMA-SC [10]	0.9094	0.9094	0.9015	0.9147	0.9457	0.9356	0.9145	0.9568	0.9154	0.9058
PRDT-DRU [2]	0.9214	0.9297	0.9245	0.9245	0.9358	0.9205	0.9245	0.9289	0.9265	0.9214
J-SSO [5]	0.9154	0.9265	0.9145	0.9234	0.9245	0.9241	0.9235	0.9265	0.9287	0.9301
Improved honeypot [6]	0.9254	0.9265	0.9278	0.9285	0.9278	0.9298	0.9314	0.9365	0.9345	0.9354
SECHS [9]	0.9254	0.9265	0.9278	0.9285	0.9278	0.9298	0.93140	0.9325	0.9345	0.9354
GMGW [1]	0.9325	0.9331	0.9339	0.9385	0.9401	0.9412	0.9423	0.9431	0.9435	0.9439
Proposed	0.9614	0.9678	0.9345	0.9874	0.9678	0.9865	0.9758	0.9741	0.9714	0.9785

Table 9 Comparison of the data restoration effectiveness with the state-of-the-art techniques using dataset-II

Techniques	Number of instances taken from the input dataset									
	1	2	3	4	5	6	7	8	9	10
SKMA-SC [10]	0.914	0.918	0.9235	0.9245	0.9269	0.9278	0.9287	0.9301	0.9310	0.9324
PRDT-DRU [2]	0.9185	0.9214	0.9245	0.9237	0.9314	0.9345	0.9359	0.9365	0.9378	0.9387
J-SSO [5]	0.9210	0.9254	0.9278	0.9284	0.9291	0.9301	0.9335	0.9401	0.9412	0.9421
Improved honeypot [6]	0.9254	0.9265	0.9301	0.9312	0.9328	0.9330	0.9345	0.9367	0.9370	0.9385
SECHS [9]	0.9354	0.9365	0.9378	0.9385	0.9391	0.9398	0.9399	0.9401	0.9402	0.9412
GMGW [1]	0.9401	0.9406	0.9412	0.9423	0.9429	0.9430	0.9435	0.9438	0.9441	0.9454
Proposed	0.9632	0.9701	0.724	0.9789	0.9814	0.9826	0.9834	0.9845	0.9854	0.9867

Table 10 Comparison of the data restoration effectiveness with the state-of-the-art techniques using dataset-III

Techniques	Number of instances taken from the input dataset									
	1	2	3	4	5	6	7	8	9	10
SKMA-SC [10]	0.9000	0.9002	0.9007	0.9014	0.9025	0.9039	0.9089	0.9105	0.9124	0.9174
PRDT-DRU [2]	0.9101	0.9129	0.9135	0.9145	0.9196	0.9205	0.9215	0.9219	0.9229	0.9236
J-SSO [5]	0.9106	0.9124	0.9136	0.9149	0.9293	0.9301	0.9305	0.9314	0.9326	0.9327
Improved honeypot [6]	0.9212	0.9222	0.9238	0.9249	0.9257	0.9260	0.9278	0.9297	0.9301	0.9374
SECHS [9]	0.9260	0.9265	0.9278	0.9285	0.9301	0.9325	0.9333	0.9389	0.9401	0.9421
GMGW [1]	0.925	0.9271	0.9281	0.9289	0.9294	0.9305	0.9312	0.9324	0.9378	0.9389
Proposed	0.958	0.9612	0.9635	0.9678	0.9681	0.9689	0.9695	0.9701	0.9709	0.9712

such as SKMA-SC, PRDT-DRU, J-SSO, Improved honeypot, and SECHS techniques. For the dataset-III, the proposed model performance is 4.2% higher when compared to the conventional techniques. The performance outcomes of the three datasets using our proposed methodology show that our model's restoration process is beneficial when compared to the conventional techniques.

6 Conclusion

This paper presents a Gaussian Mutation based Firebug optimization (GM-FBO) algorithm for improving the security in cloud-based healthcare applications. The main aim of this paper is to prevent the privacy of the user's medical information stored in the cloud storage server. The GM-FBO algorithm is used to offer secure data storage and restoration operations. The experiments are conducted using three different healthcare datasets, namely HPD, Medical MIMIC-III, and MHEALTH. The efficiency of the proposed model is compared with different state-of-the-art techniques such as SKMA-SC, PRDT-DRU, J-SSO, Improved honeypot, and SECHS. The convergence efficiency of the proposed GM-FBO algorithm is analysed by comparing it with different heuristic algorithms such as GA, PSO, FA, and WOA. The experiments are conducted to verify the restoration effectiveness, statistical significance, and proficiency of the proposed technique towards KPA and CPA techniques. The privacy and fitness measures are evaluated by varying the key size from 100 to 500. The simulation results show that the GM-FBO algorithm offers maximum privacy with minimal fitness value. Hence the proposed model is termed to be effective when securing the sensitive data stored by the medical institutions in the cloud. In the future, we plan to extend this work to handle different attacks such as impersonation, phishing, and botnet that take place in IoT-based healthcare applications.

Data availability Data sharing is not applicable to this article as no new data were created or analysed in this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed consent Informed consent does not apply as this was a retrospective review with no identifying patient information.

References

1. Elmisery AM, Rho S, Aborizka M (2019) A new computing environment for collective privacy protection from constrained healthcare devices to IoT cloud services. Clust Comput 22(1):1611–1638

2. Sathya A, Raja SKS (2021) Privacy preservation-based access control intelligence for cloud data storage in smart healthcare infrastructure. Wirel Pers Commun 118(4):3595–3614
3. Sendhil R and Amuthan A (2021) Contextual fully homomorphic encryption schemes-based privacy preserving framework for securing fog-assisted healthcare data exchanging applications. Int J Inform Technol, pp.1–9
4. Liu H, Yao X, Yang T, Ning H (2018) Cooperative privacy preservation for wearable devices in hybrid computing-based smart health. IEEE Internet Things J 6(2):1352–1362
5. Xie Q, Sundararaj V, Mr R (2021) Analyzing the factors affecting the attitude of public toward lockdown, institutional trust, and civic engagement activities. J Community Psychol. https://doi.org/10.1002/jcop.22681
6. Mewada S, Gautam SS, Sharma P (2020) Artificial bee colony-based approach for privacy preservation of medical data. Int J Inform Syst Mod Des (IJISMD) 11(3):22–39
7. Madan S, Goswami P (2020) A privacy preservation model for big data in map-reduced framework based on k-anonymisation and swarm-based algorithms. Int J Intell Eng Inform 8(1):38–53
8. Kalia P, Bansal D and Sofat S (2021) Privacy preservation in cloud computing using randomized encoding. Wirel Pers Commun, pp.1–13
9. Tu NA, Wong KS, Demirci MF, Lee YK (2021) Toward efficient and intelligent video analytics with visual privacy protection for large-scale surveillance. J Supercomput. https://doi.org/10.1007/s11227-021-03865-7
10. Mansour HO, Siraj MM, Ghaleb FA, Saeed F, Alkhammash EH, Maarof MA (2021) Quasi-Identifier recognition algorithm for privacy preservation of cloud data based on risk reidentification. Wirel Commun Mob Comput 2021:7154705. https://doi.org/10.1155/2021/7154705
11. Alphonsa MA, Amudhavalli P (2018) Genetically modified glowworm swarm optimization based privacy preservation in cloud computing for healthcare sector. Evol Intel 11(1):101–116
12. Wu X, Xu X, Dai F, Gao J, Ji G, and Qi L (2020) An Ensemble of Random Decision Trees with Personalized Privacy Preservation in Edge-Cloud Computing. In: 2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), pp. 779–786. IEEE
13. Abdo MA, Abdel-Hamid AA. and Elzouka HA (2020) A Cloud-based Mobile Healthcare Monitoring Framework with Location Privacy Preservation. In: 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT) (pp. 1–8). IEEE
14. Khatiwada P, Bhusal H, Chatterjee A and Gerdes MW (2020) A Proposed Access Control-Based Privacy Preservation Model to Share Healthcare Data in Cloud. In: 2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)(50308) (pp. 40–47). IEEE
15. Ahamad D, Hameed SA, Akhtar M (2020) A multi-objective privacy preservation model for cloud security using hybrid Jaya-based shark smell optimization. J King Saud Univ Comput Inform Sci. https://doi.org/10.1016/j.jksuci.2020.10.015
16. Mondal A, Goswami RT (2021) Enhanced Honeypot cryptographic scheme and privacy preservation for an effective prediction in cloud security. Microprocess Microsyst 81:103719
17. Yang X, Wang M, Wang X, Chen G, Wang C (2020) Stateless cloud auditing scheme for non-manager dynamic group data with privacy preservation. IEEE Access 8:212888–212903
18. Fang C, Guo Y, Wang N, Ju A (2020) Highly efficient federated learning with strong privacy preservation in cloud computing. Comput Secur 96:101889
19. Jayaram R, Prabakaran S (2020) Onboard disease prediction and rehabilitation monitoring on secure edge-cloud integrated privacy preserving healthcare system. Egypt Inform J. https://doi.org/10.1016/j.eij.2020.12.003
20. Prabha KM, Saraswathi PV (2020) Suppressed K-anonymity multi-factor authentication based schmidt-samoa cryptography for privacy preserved data access in cloud computing. Comput Commun 158:85–94
21. Mandala J, Rao MCS (2019) Privacy preservation of data using crow search with adaptive awareness probability. J Inform Secur Appl 44:157–169
22. Noel MM, Muthiah-Nakarajan V, Amali GB and Trivedi AS (2021) A new biologically inspired global optimization algorithm based on firebug reproductive swarming behaviour. Exp Syst Appl, p.115408
23. Karthik E and Sethukarasi T (2021) Sarcastic user behavior classification and prediction from social media data using firebug swarm optimization-based long short-term memory. The J Supercomput, pp.1–25

24. Sulis E, Terna P, Di Leva A, Boella G, Boccuzzi A (2020) Agent-oriented decision support system for business processes management with genetic algorithm optimization: an application in healthcare. J Med Syst 44(9):1–7
25. Chen H, Zhang Q, Luo J, Xu Y, Zhang X (2020) An enhanced bacterial foraging optimization and its application for training kernel extreme learning machine. Appl Soft Comput 86:105884
26. Shivashankar M, Mary SA (2021) Privacy preservation of data using modified rider optimization algorithm: optimal data sanitization and restoration model. Exp Syst 38(3):e12663
27. Asaumya (2021) HealthCare Problem: Prediction Stroke Patients, Kaggle, 17-Jul-2018. [Online]. Available: https://www.kaggle.com/asaumya/healthcare-problem-prediction-stroke-patients. [Accessed: 20-Oct-2021]
28. Johnson AE, Pollard TJ, Shen L, Li-Wei HL, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, Mark RG (2016) MIMIC-III, a freely accessible critical care database. Sci data 3(1):1–9
29. Banos O, Garcia R, Holgado-Terriza JA, Damas M, Pomares H, Rojas I, Saez A and Villalonga C (2014) mHealthDroid: a novel framework for agile development of mobile health applications. In: International workshop on ambient assisted living. Springer, Cham pp. 91–98
30 Bansal JC (2019) Particle swarm optimization. Evolutionary and swarm intelligence algorithms. Springer, Cham, pp 11–23
31. Yang XS (2009) Firefly algorithms for multimodal optimization. In: International symposium on stochastic algorithms. Springer: Berlin, Heidelberg pp. 169–178
32. Mirjalili S, Lewis A (2016) The whale optimization algorithm. Adv Eng Softw 95:51–67
33. Miranda-López V, Tchernykh A, Babenko M, Avetisyan A, Toporkov V, Drozdov AY (2020) 2Lbp-RRNS: two-levels RRNS with backpropagation for increased reliability and privacy-preserving of secure multi-clouds data storage. IEEE Access 8:199424–199439
34. Sundararaj V, Muthukumar S, Kumar RS (2018) An optimal cluster formation based energy efficient dynamic scheduling hybrid MAC protocol for heavy traffic load in wireless sensor networks. Comput Secur 77:277–288
35. Sundararaj V (2016) An efficient threshold prediction scheme for wavelet based ECG signal noise reduction using variable step size firefly algorithm. Int J Intell Eng Syst 9(3):117–126
36. Sundararaj V (2019) Optimised denoising scheme via opposition-based self-adaptive learning PSO algorithm for wavelet-based ECG signal noise reduction. Int J Biomed Eng Technol 31(4):325
37. Sundararaj V, Anoop V, Dixit P, Arjaria A, Chourasia U, Bhambri P, MR, Rejeesh. and Regu Sundararaj, (2020) CCGPA-MPPT: Cauchy preferential crossover-based global pollination algorithm for MPPT in photovoltaic system. Prog Photovoltaics Res Appl 28(11):1128–1145
38. Ravikumar S, Kavitha D (2021) CNN-OHGS: CNN-oppositional-based Henry gas solubility optimization model for autonomous vehicle control system. J Field Robot. https://doi.org/10.1002/rob.22020
39. Ravikumar S and Kavitha D (2020) IoT based home monitoring system with secure data storage by Keccak–Chaotic sequence in cloud server. J Ambient Intell Human Comput pp.1–13
40. Rejeesh MR (2019) Interest point based face recognition using adaptive neuro fuzzy inference system. Multimed Tools Appl 78(16):22691–22710
41. Kavitha D, Ravikumar S (2021) IOT and context-aware learning-based optimal neural network model for real-time health monitoring. Trans Emerg Telecommun Technol 32(1):e4132
42. Edwin AC, Madheswari AN (2013) Job scheduling and VM provisioning in clouds. In: 2013 Third International Conference on Advances in Computing and Communications. IEEE, pp 261–264
43. Nirmal Kumar SJ, Ravimaran S, Alam MM (2020) An effective non-commutative encryption approach with optimized genetic algorithm for ensuring data protection in cloud computing. Comput Model Eng Sci 125(2):671–697
44. Gowthul Alam MM, Baulkani S (2017) Reformulated query-based document retrieval using optimised kernel fuzzy clustering algorithm. Int J Bus Intell Data Min 12(3):299
45. Alam MG, Baulkani S (2016) A hybrid approach for web document clustering using K-means and artificial bee colony algorithm. Int J Intell Eng Syst 9(4):11–20
46. Madheswari AN (2013) Performance optimized routing for SLA enforcement in cloud computing. In: 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE). IEEE, pp 689–693
47. Sundararaj V, Selvi M (2021) Opposition grasshopper optimizer based multimedia data distribution using user evaluation strategy. Multimed Tools Appl 80(19):29875–29891

48. Rejeesh MR, Thejaswini P (2020) MOTF: multi-objective optimal trilateral filtering based partial moving frame algorithm for image denoising. Multimed Tools Appl 79(37):28411–28430

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