Peer-to-Peer Networking and Applications
https://doi.org/10.1007/s12083-022-01322-7

# An enhanced bacterial foraging optimization algorithm for secure data storage and privacy-preserving in cloud

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

Received: 23 March 2021 / Accepted: 3 April 2022

**Abstract**
Cloud file access is the most widely used peer-to-peer (P2P) application, in which users share their data and other users can access it via P2P networks. The need for security in the cloud system grows day by day, as organizations collect a massive amount of users' confidential information. Both the outsourced data and the unprotected user's sensitive data need to be protected under the cloud security claims since the advanced P2P networks are prone to damage. The recurring security breach in the cloud necessitates the establishment of an advanced legal data protection strategy. Various researchers have attempted to develop privacy-preserving cloud computing systems employing Artificial Intelligence (AI) techniques, however, they have not been successful in achieving optimal privacy. AI approaches implemented in the cloud assist applications in efficient data management by analyzing, updating, classifying, and providing users with real-time decision-making support. AI approaches can also detect fraudulent activity by analyzing deviations in normal data patterns entering the system. To handle the security concerns in the cloud, this paper presents a novel cybersecurity architecture using the Chaotic chemotaxis and Gaussian mutation-based Bacterial Foraging Optimization with a genetic crossover operation (CGBFO-GC) algorithm. The CGBF0-GC algorithm cleanses and restores the data using a multiobjective optimal key generation mechanism based on the following constraints: data preservation, modification, and hiding ratio. The simulation results show that the proposed methodology outperforms existing methods in terms of convergence, key sensitivity analysis, and resistance to known and chosen-plaintext attacks.

**Keywords** Cloud computing · Security · CGBFO- GC algorithm · Data cleansing · And data restoration

## 1 Introduction

In a cloud computing environment, all resources and applications are provided through on-demand platforms that are available through the internet. Several potential cloud customers are reluctant to use cloud computing because of privacy and security concerns. The security feature of some cloud computing communications [1] is a critical component. Generally, the infrastructure as services, platform as service, and software as service are provided by the cloud. When data is saved in a data center, it raises concerns over maintaining customer data security. In a range of disciplines, such as business, medical, and defense, the cloud computing industry contributes significantly to the global environment [2]. Data confidentiality is one of the significant features of a cloud environment.

User and physical access control, audit scheduling, identity and access management, encryption, and key management are among the security issues identified in the cloud data. The privacy features of a cloud environment are based on securing the data storage and processing. Both cloud consumers and cloud servers have the same level of cloud security, and trust is a must [3]. In recent years, a variety of encryption methods have been used to encrypt data. In the cloud sector, privacy-preserving approaches are used to achieve different objectives. The fuzzy Grouping attribute was presented to provide a privacy-aware access control strategy for improved

✉ K. Anand
anandonmail@gmail.com; anandk@citchennai.net

[1] Center for Artificial Intelligence and Research, Chennai Institute of Technology, Chennai, Tamil Nadu, India

[2] Department of Information Technology, Vignan's Foundation for Science Technology and Research (Deemed to be University), Vadlamudi, Guntur, Andhra Pradesh- 522213, India

[3] Department of Computer Science and Engineering, Saveetha Engineering College(Autonomous), Chennai, Tamil Nadu, India

Published online: 28 May 2022

data privacy in the cloud sector. This verification scheme's efficiency is still inefficient [4].

The user security and privacy in the cloud environment are protected using the Attribute-Based Signature Outsourcing (ABSO) protocol but it has extremely higher computational complexities. In order to provide security, the perturbation-oriented model modifies the noise data[5]. The careful calibration, on the other hand, necessitated a change to locate the usability for improved stability and model privacy. When it comes to plaintext, privacy attributes are linked to security concerns. In the cloud, a large amount of individual data is generated, and the CSPs analyze the ownership and responsibilities associated with the data. Sensitive data handling is simpler if the CSP is considered to be confidential [6]. The above-mentioned problems are solved using numerous approaches in the last decades [7–19]. Hence, we needed an effective cloud security model to overcome these issues present in the existing techniques. The high cost associated with the centralized cloud servers is overcome in this approach using a Peer to Peer (P2P) cloud storage. The P2P secure cloud infrastructure offers efficient web service hosting and data storage. In this study, we have proposed data sensitization and restoration with CGBFO- GC algorithm for optimal key generation in the cloud data security model. The major contributions of this paper are summarized as shown below:

- The novel CGBFO-GC algorithm is formulated by integrating the Genetic Crossover Operation based Bacterial Foraging Optimization algorithm along with the Chaotic Chemotaxis and Gaussian Mutation method to propose an optimal key generation approach.
- Multi-objective functions including data preservation ratio, hiding ratio, and modification degree are satisfied using the CGBFO-GC algorithm to offer secure cloud storage.
- The proposed method demonstrates optimal and superior results in terms of key sensitivity evaluation, CPA and KPA attack analysis, and Convergence analysis.

The rest of the paper is arranged as: Sect. 2 explains the existing methods based on the cloud security model. Section 3 delineates the system architecture. The proposed methodology concerning optimal key generation using the proposed CGBFO- GC algorithm is delineated in Sect. 4 followed by the simulation results in Sect. 5. At last, Sect. 6 concludes the paper.

## 2 Review of related works

Artificial intelligence methodologies were developed by Ahamad et al. [20] to design a cloud-based privacy replica. Cloud computing provides cost savings, and agility for businesses to offer high flexibility by hosting the data. The two main steps of their privacy preservation system are data restoration and sanitization. According to optimal key generation, the hybrid meta-heuristic algorithm named Jaya-based Shark Smell Optimization (J-SSO) performs the sanitization process. Optimal key generation is achieved by deriving multi-objective functions such as data preservation ratio, modification degree, and hiding ratio. This method provided less space but with a higher computational cost. A Reputation-aware Trust and Privacy Preservation (RTPP) scheme was introduced by Ahmad et al. [21] for mobile cloud computing. Trust management is addressed via the usage of cloud services and the selection of CCs based on reputation. For security solutions and key management, a hybrid policy tree mechanism is proposed to select a dynamic attribute. The AS-CABE efficiency against security attacks is analyzed to present security analysis. The RTPP with AS-CABE demonstrated higher performances in the case of resilience, trust, computation, storage, reputation score, encryption, and decryption time but the execution time is higher.

Li et al. [22] developed non-abelian rings for homomorphic encryption via matrix-ring. The intermediate result of any ciphertext operations is a fast ciphertext homomorphic comparison without decryption. According to the encryption strategy, the maximum ciphertext expansion rate required more efficiency. Abirami et al. [23] demonstrated improved cloud security by using a crypto-deep neural network in a trusted environment to maintain anonymity. The crypto-deep neural network enhances distributed security. This model consists of a cloud agent, data center, web server, and cloud server. Impersonation attacks are objectives for crypto-deep neural network cloud security (CDNNCS). The linear algebraic equation scheme is secured by improving the level of trust between cloud users. The CDNNCS accomplished superior performances in terms of throughput, Jitter, and Delay with minimum packet loss but required a more optimal traffic and security model.

Park et al. [24] proposed a security vulnerability measurement with cost-optimization for efficient security enhancement. Vulnerability-based mitigation and risk assessment are used to improve security measures while working with a constrained security budget. According to security vulnerability measurement, the security cost allocation strategies are evaluated. This model required low-latency resources, data analysis, and various environmental results. For resource allocation with security concerns, Meng et al. [25] suggested security-aware scheduling based on distributed Particle Swarm Optimization (SAS-DPSO). The dynamic workflow model based on a dynamic scheduling mechanism is introduced by means of the mobile industrial application mobility and dynamics of edge resources. The experiment's findings were effective in striking an effective balance between security and scheduling speed, despite requiring higher latency resources and having limited scalability.

## 3 System model and Architecture

In the current research, cloud computing security is a significant factor to be addressed. If the security measures for data transmission and operation are not given correctly, the data will be at risk. Organizations usually design applications based on decentralized P2P cloud architecture and allocate a higher budget to security because they are vulnerable to attacks. As cloud storage provides the ability, the possibility of an elevated risk in data dispensation and the stored data is examined by a number of users. To deal with cloud security constraints, the security challenge and solutions are identified through establishing appropriate security measures. When large organizations share resources, there is a risk of data misuse [26]. As a result, data and data archives must be safeguarded to minimize risks. To overcome the constraints of current data security techniques in the literature, we provide a novel cybersecurity paradigm for cloud data. The decentralized P2P cloud architecture's cybersecurity paradigm protects users from remote exploits and creates a secure network protocol that prevents malicious or suspicious services from invading.

The proposed cybersecurity model is evaluated using the datasets obtained from UCI repositories such as Wholesale customer data (Dataset-1), Heart disease (Dataset-2), and Air quality data (Dataset-3). Data cleansing and data restoration are two major stages of the projected privacy conservation scheme. In a cloud, removing the frequent data errors such as missing values and typos is called data cleansing. As a result, exposure to an unauthorized point is avoided. The proposed CGBFO- GC, on the other hand, is used for data cleansing [27] via optimal key generation. The use of a multiobjective function regularises optimal key generation by taking into account different criteria such as data modification degree, hiding ratio, and preservation ratio. The data cleansing and restoration in the cloud are carried out utilizing this multi-objective function with CGBFO- GC algorithm.

## 4 Proposed approach

In this section, we discuss data cleansing, data restoration, and the optimal key generation using the CGBFO-GC algorithm (Fig. 1).

### 4.1 Dataset details

Three datasets from the UCI repository related to cloud data were used in this study and the detailed description of each dataset is provided as follows:

***Wholesale customer data:*** Eight attributes with 440 instances present in the wholesale customer dataset (https://archive.ics.uci.edu/ml/datasets/Wholesale+customers) [28]. The Lisbon, Oporto, region, deter gents-paper, delicatessen, milk, frozen, grocery, fresh, etc. are examples of the few attributes.

***Heart disease data:*** There are 303 instances with 75 attributes included in the heart disease dataset(https://archive.ics.uci.edu/ml/datasets/Heart+Disease) [29]. This dataset considers id, age, chol, htn, smoke, and cigs.

***Air quality data:*** The air quality dataset (https://archive.ics.uci.edu/ml/datasets/Air+Quality) contains 9358 instances from five metal oxide chemical sensor arrays [30].

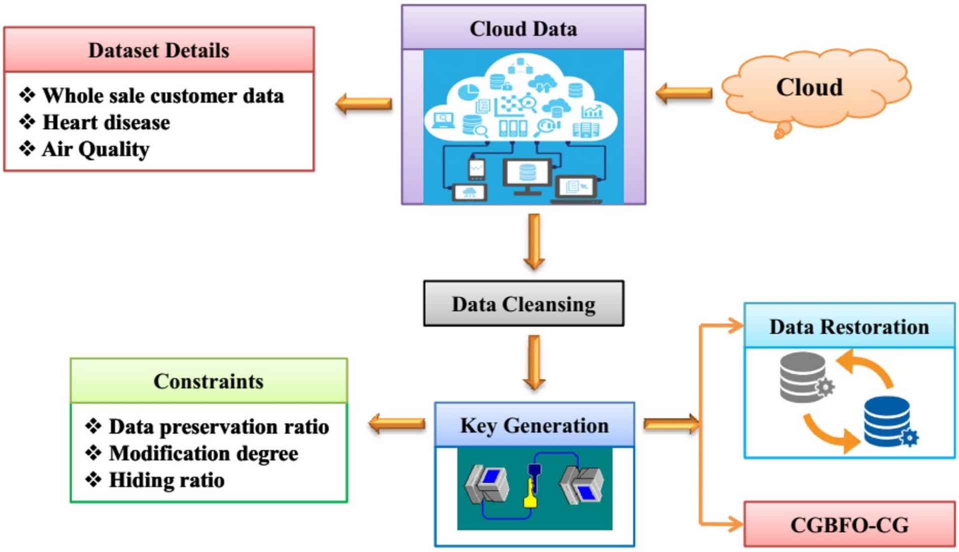


**Fig. 1** Proposed privacy preservation architectural diagram

The sensor is located in the most polluted region of the Italian city. From March 2004 to February 2005, data with various attributes such as relative humidity, data, temperature, absolute humidity, reference analyzer, and time is chosen.

## 4.2 Dataset cleansing and restoration

Hiding sensitive information or data in a cloud process is data cleansing. Data is prevented from escaping to an unauthorized point. Data cleansing is the reverse operation of data restoration. Figure 2 illustrates the data cleansing and restoration process. The key matrix generation and cloud data determine the binary conversion during cleansing. The optimal key is generated using the CGBFO-GC algorithm. Using the XOR technique, the cleansing data extracts received binary data. Equation (1) produces the cleansing data from the key matrix generation and original cloud data.

$$C'_{data} = C_{data} \oplus Key_2 \tag{1}$$

where $\underset{data}{\overset{\wedge}{C}}$ is the cleansing data and $C_{data}$ is the original data. Here, $Key_2$ is an optimally generated key. The cleansing is performed using $C'_{data}$, after cleansing, $C_{data}$ accomplished is known as the intentionality of the proposed representation. The cleansing process hides sensitive rules and transfers them to the cloud. The security of the cloud division has been enhanced, and the data has been isolated for future usage. The proposed CGBFO-GC algorithm with the same key recovers the original data during restoration. Equation (2) explains the restoration process.

$$\underset{data}{\overset{\wedge}{C}} = C'_{data} \oplus Key_2 \tag{2}$$

where the restored data is $\underset{data}{\overset{\wedge}{C}}$.

## 4.3 Proposed Optimal Key Generation Mechanism

The major role in cleaning and restoration is key extraction in the proposed cloud data cybersecurity model. In this section, we used the CGBFO-GC algorithm for optimal key generation. The Kronecker method converts the key into a new form. Equation (3) converts the key into $Key_1$. Where $\sqrt{Nu^N} \times Max_p$ considers the size for both $Key_1$ and $Key_2$. The key matrix for the $Key = 5, 6, 1$ is generated using the below equation:

$$Key_1 = \begin{bmatrix} 5 & 5 & 5 \\ 6 & 6 & 6 \\ 1 & 1 & 1 \end{bmatrix}_{\left[\sqrt{Nu^N} \times Max_p\right]} \tag{3}$$

where, $Nu$ is the number of transactions with the highest score $Nu'$ and $Max_p$ is the maximum transaction length. The reconstructed key matrix $Key_1$ is created via row-wise duplication. The Kronecker method generates the key matrix $Key_2$ as shown in Eq. (4).

$$Key_2 = key_1 \oplus key_1 \tag{4}$$

The Kronecker product is represented using a symbol $\oplus$. The proposed CGBFO-GC method is used for best key making in the cloud security model. The steps of the CGBFO-GC algorithm for optimal key generation are as follows:

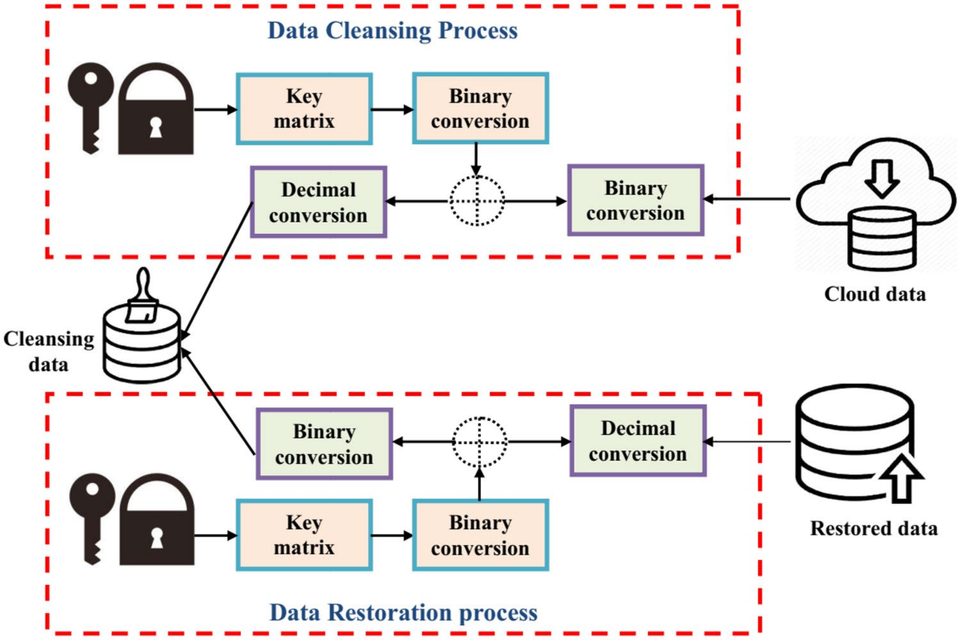


**Fig. 2** Data cleansing and restoration process

#### 4.3.1 Operation of chaotic chemotaxis step length

The Bacterial Foraging Optimization (BFO) algorithm to change the original set chemotaxis step length [30]. The Gaussian mutation is followed by a genetic crossover operation (GCO) in the chemotaxis step. The optima positions in the current bacterial individuals are operated by allowing chemotaxis step. Equation (5) introduces the chaos theory as shown below:

$$CH_{l+1} = \gamma CH_l * (1 - CH_l) \qquad l = 1, \ldots\ldots, R - 1 \tag{5}$$

where the control parameter is $\gamma$. The bacterial population is $R$. The logistic and random maps are used to locate the intervals 0 and 1. The logistic map generates the chaotic series $CH$. The logistic map contains a higher probability of creating standards near zero and one than the random map between the intervals 0 and 1. The chaotic search is very effective in local optimization owing to its randomicity and ergodicity [31].

The obtained chaotic chemotaxis step length is sorted using Eq. (6). The falling into local optimal predicament is prevented via chemotaxis step length.

$$D = sort\ (D', descend') \tag{6}$$

#### 4.3.2 Gaussian mutation process

The mutated position $G_bGao$ is generated. According to the current bacterial population, the Gaussian mutation is applied towards the best position $G_b$.

$$G_bGoa = G_b * (1 + Gaus(0, 1)) \tag{7}$$

The standard normal distribution is $Gaus(0, 1)$. The $G_b$ value is replaced with the fitness of $G_bGao$ if the mutated position of the fitness functions $G_bGao$ is better than $G_b$ then update $G_b$ with $G_bGao$. Based on $G_b$, the Gaussian distributed random disturbance term $G_b * Gaus(0, 1)$ increased using Eq. (8). The convergence speed is improved that converges to global optima.

#### 4.3.3 Genetic crossover operation (GC)

The key part of the basis bat algorithm (BA) is local searching. For complex tasks, the random walk is enough. The Genetic Algorithm (GA) with BA combination crossover operation is used to solve this problem [31]. The crossover technique [32] is used to develop the novel solution among the most optimal solution and the global optimal solution of the current iteration. Equation (8) defines the mathematical representation of the crossover operation.

$$\overline{Y_{new}} = PG_b * (1 - E) + Y_b * E \tag{8}$$

The current iteration with the best solution is $PG_b$ and the global best solution is $Y_b$. Where the steady-state is $E$. When the global best solution is unequal towards the most excellent solution of the current iteration then the operation of genetic crossover is preceded [33]. Figure 3 explains the overall procedure of the CGBFO-GC algorithm.

### 4.4 Cloud cyber security-based objective model:

In this work, the three major objective functions including Modification degree, Hiding ratio, and Data preservation ratio are delineated as follows:

#### 4.4.1 Modification degree

The modification degree describes the modification level that occurs between the cleansed dataset $C^{\}}_{data}$ and the original dataset $C_{data}$ in which the Euclidean distance determination is measured. Equation (9) formulates the modification degree [34].

$$G_1 = C_{data} - C'_{data} \tag{9}$$

#### 4.4.2 Hiding ratio

The sensitive rate defines the hiding ratio in which is correctly concealed in $C'_{data}$. The difference among the original data of the relevant index considers the term $E_1$. Equation (10) offers the difference between $E_1$ and $E_2$.

$$E_{difference} = xyz(E_1 - E_2) \tag{10}$$

The non-zero indexes of $E_{difference}$ length are $LN_1$. Equation (10) explains the mathematical form of the hiding ratio.

$$H_{iding}R_{atio} = \frac{LN_1}{H_{data}} \tag{11}$$

where, $H_{data}$ is the number of data indexes have to hide. The best performance maximizes the hiding ratio.

#### 4.4.3 Ratio of data preservation

The non-sensitive rules rate defines the information preservation ratio that conceals $H_{data}$ [35]. Equation (12) represents the ratio of data preservation.

$$D_{preservation} = \frac{LN_2}{P_{data}} \tag{12}$$

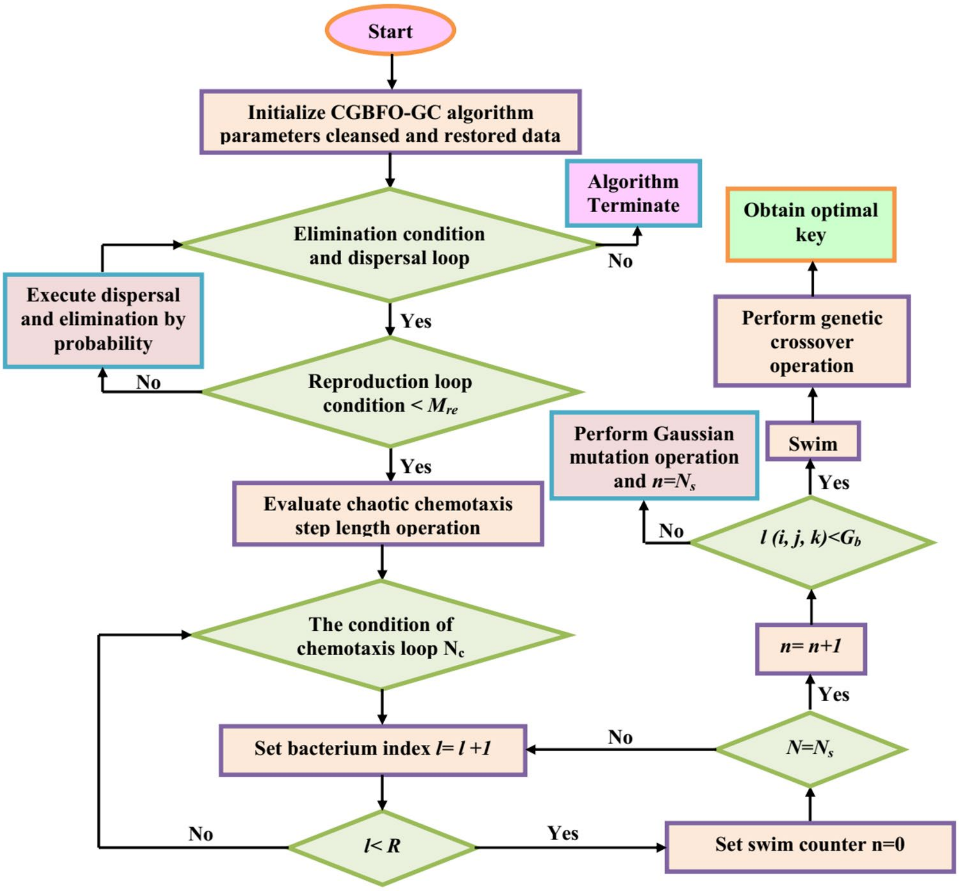

Start
Initialize CGBFO-GC algorithm parameters cleansed and restored data
Elimination condition and dispersal loop
No
Algorithm Terminate
Obtain optimal key
Execute dispersal and elimination by probability
Yes
Perform genetic crossover operation
No
Reproduction loop condition < $M_{re}$
Yes
Perform Gaussian mutation operation and $n=N_s$
Swim
Yes
Evaluate chaotic chemotaxis step length operation
No
$l\ (i, j, k)<G_b$
The condition of chemotaxis loop $N_c$
$n= n+1$
Yes
Set bacterium index $l= l+1$
No
$N=N_s$
No
$l< R$
Yes
Set swim counter n=0

**Fig. 3** Working of the CGBFO-GC algorithm

where, $LN_2$ is the number of zero indexes and the term $P_{data}$ preserving the total number of data indexes. The proposed security model maximizes the preservation ratio.

### 4.4.4 Encode the solution

The data cleansing and restoration are performed by using the proposed CGBFO-GC algorithm for key optimization. The key length is changed according to the number of transactions or data size. Where, $N_l = 1, 2, ...., N_o$. From this, the attribute or field length is $N_o$. From *1* to $2^6$–*1*, that gives the bounding limit. To generate the best solution, the CGBFO-GC algorithm optimizes the key vectors [36].

### 4.4.5 Last objective function

The multi-objective function achieves optimal key generation. Equation (13) represents the parameters such as data preservation, hiding ratio, and modification degree [37].

$$Obj = H_1 + (1 - GS) + (1 - HS) \tag{13}$$

where the constant terms are *GS* and *HS*. The data preservation, hiding ratio, and modification degree are $H_1$.

## 5 Result and Discussion

The performance of the proposed work is discussed using various performance metrics with a state-of-art comparison. MATLAB 2018 is a software that implements the model of proposed cloud data cybersecurity. As one of the features of MATLAB, virtualization allows for extensive data analysis and the development of computational algorithms. The complete solution is constructed to leverage different technologies given in MATLAB to handle the cloud data. The size of the population is 10 and the maximum numeral of iteration is 50 in the experiment [38, 39]. Table 1 explains the parameter settings of the proposed method. Key sensitivity analysis and state-of-the-art comparison are used to evaluate the proposed method's efficiency. In this experiment, we have chosen SAS-DPSO, CDNNCS, J-SSO [20], and GC [33] with the proposed CGBFO-GC algorithm as the state-of-art method. The relevant parameters including hiding ratio, modification degree, and the ratio of data preservation are considered to obtain the best key making using the proposed CGBFO-GC method. The loss of information is visible in the cleansed data when compared to the original information and the degree of modification [40]. The aim of not hiding other data and hiding sensitive data, according to

**Table 1** Parameter settings based on proposed CGBFO-GC algorithm

| Parameters | Ranges |
|---|---|
| Population size | 10 |
| Maximum number of iteration | 50 |
| Swimming length | 5 |
| Crossover ratio | 0.9 |
| Mutation ratio | 0.1 |

data preservation, is to successfully demonstrate the hiding rate. The most effective key generation is explored, as well as data cleansing and restoration capabilities. This method verifies the securitsy of all cloud data types.

### 5.1 Modification degree analysis

Figure 4 depicts the performance of modification degree analysis using three datasets 1,2, and 3. The graph is plotted between several iterations and distances. A Euclidian distance between cleansed data and original data is the modification degree as mentioned earlier [41, 42]. While correlated along with the original data, this demonstrates the loss of information that occurred in cleansed data. During the data cleansing process, it ensures that there is no loss of information and the modification degree of the cleansed data is minimum. For all the iterations from 0 to 50, minimal distance than the conventional algorithm produced by the proposed method. The proposed CGBFO-GC

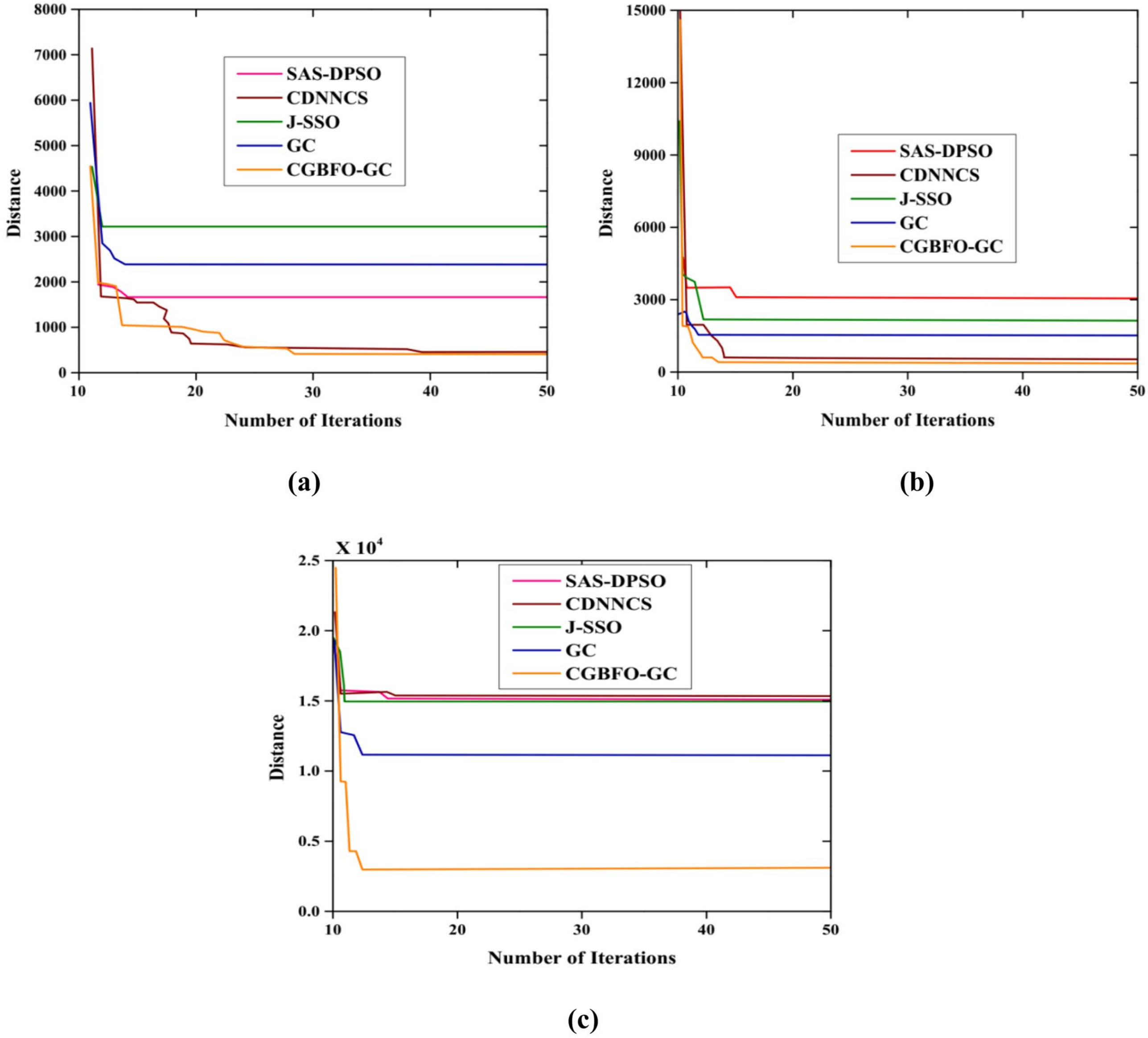


**Fig. 4** Modification degree performance analysis, (**a**) Dataset-1, (**b**) Dataset-2, and (**c**) Dataset-3

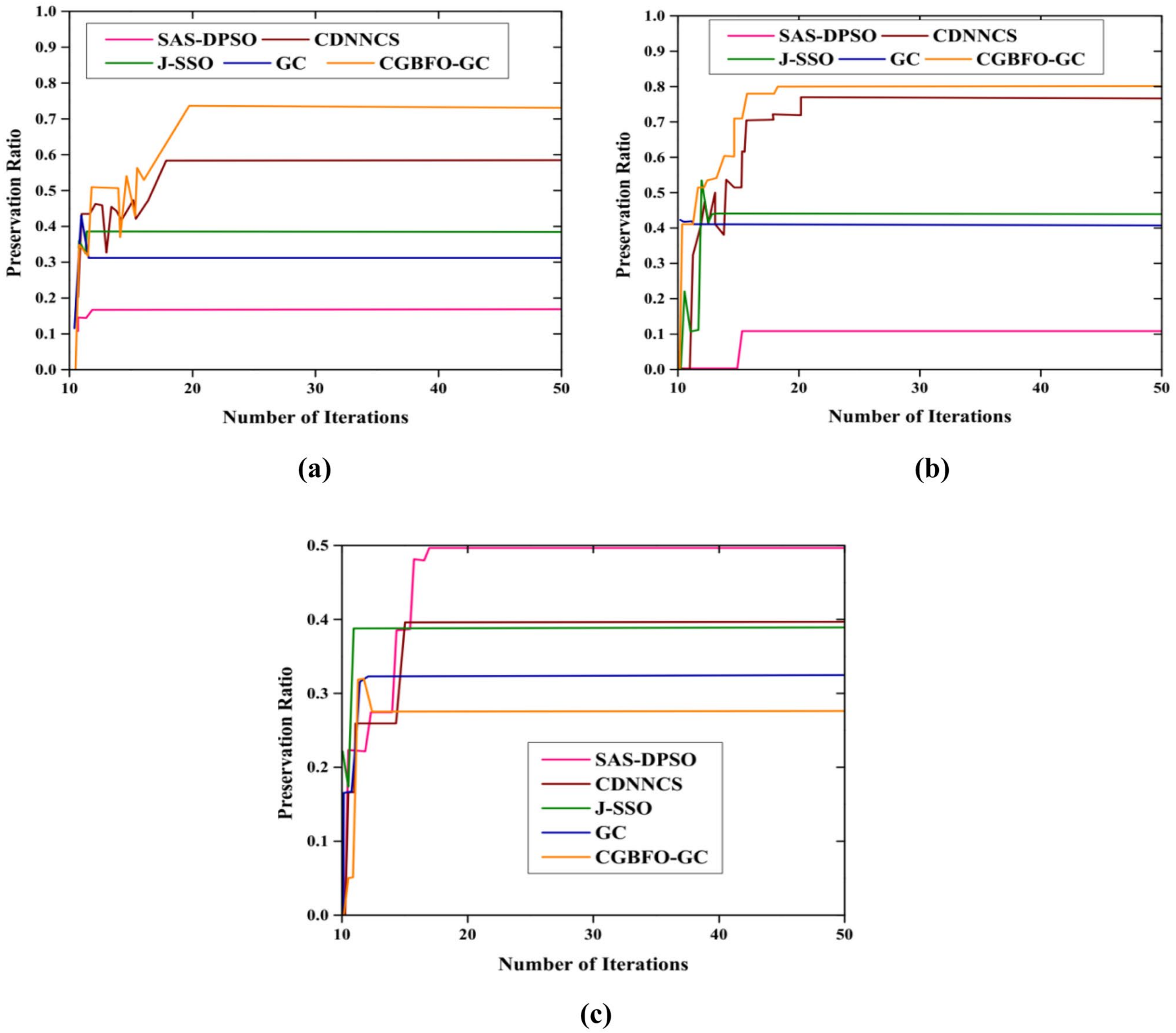
algorithm accomplished optimal modification degree results than existing SAS-DPSO, CDNNCS, J-SSO, and GC methods.

## 5.2 Preservation ratio analysis

The preservation ratio performance analysis for datasets 1, 2, and 3 are depicted in Fig. 5a–c. For all the datasets, the proposed CGBFO-GC algorithm demonstrates an enhanced ratio of preservation. For various iterations, the proposed CGBFO-GC algorithm demonstrates better results than conventional algorithms. For dataset 1, the CGBFO-GC algorithm accomplishes 0.75%, 0.85% and 2.5%. The conventional methods regarding the relevant preservation data in the cloud outperform all the datasets of the proposed CGBFO-GC algorithm.

## 5.3 Performance analysis based on convergence

Figure 6 delineates the performance of convergence of cloud security. The proposed method's convergence performance in terms of datasets 1, 2, and 3 for cloud security is demonstrated in Figs. 6a–c. The convergence value of the proposed CGBFO-GC method accomplishes 1.2%, 0.2%, and 4% values for datasets 1, 2, and 3. However, the proposed CGBFO-GC method demonstrates optimal performances than existing methods such as SAS-DPSO, CDNNCS, J-SSO, and GC methods.



(a)

(b)

(c)

**Fig. 5** Preservation ratio performance analysis, (**a**) Dataset-1, (**b**) Dataset-2, and (**c**) Dataset-3

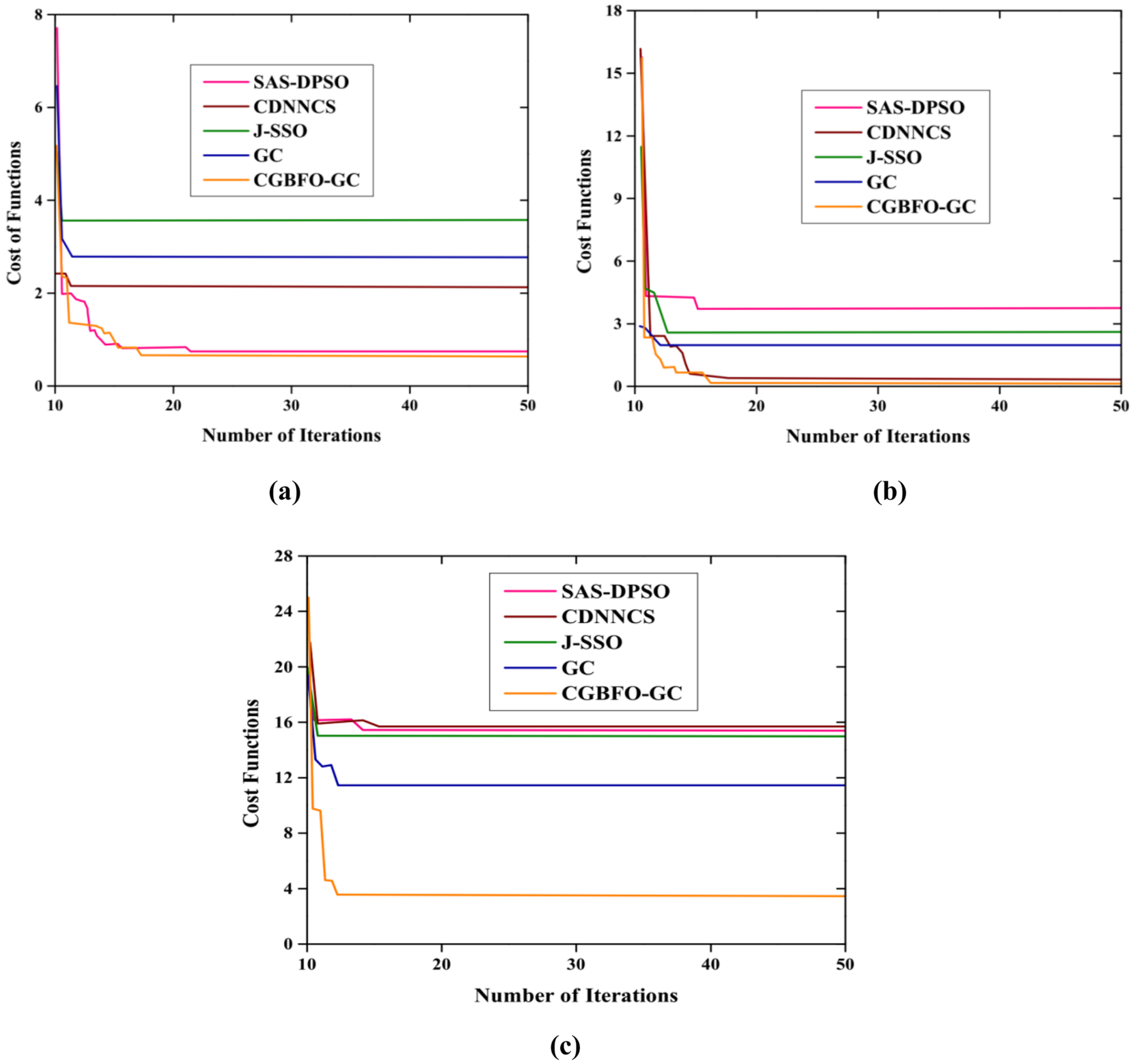


**Fig. 6** Convergence performance analysis, (**a**) Dataset-1, (**b**) Dataset-2, and (**c**) Dataset-3

## 5.4 CPA and KPA attacks analysis

For arbitrary plaintexts, the attack model describes the CPA that presumes ciphertexts are obtained through the attacker. Both encrypted version and plaintext are accessed by an attacker in which the attack model for cryptanalysis is KPA. Table 2 delineates the different attack effects using various datasets. If the CPA attack is completed then the correlation among original data and restored information is calculated in this section. The minimum connection between restored and original information is obtained when the KPA attack is performed.

***KPA attack effect analysis:*** For whole sale customer data, the SAS-DPSO, CDNNCS, J-SSO, GC and proposed CGBFO-GC methods demonstrated 0.9999%, 1%, 1%, 1% and 0.9979%. Further,SAS-DPSO, CDNNCS, J-SSO, GC and proposed CGBFO-GC methods obtained 0.9969%, 0.9983%, 0.9972%, 0.99979% and 0.9949% for heart disease data. Similarly, the SAS-DPSO, CDNNCS, J-SSO, GC and proposed CGBFO-GC methods provided 0.9998%, 0.9992%, 0.9998%, 0.9999% and 0.9997% for air quality datasets.

***CPA attack effect analysis:*** The CPA attack effect is analyzed using different state-of-art methods including DPSO,

**Table 2** Attack effect analysis using a different dataset

| **Name of the attacks** | **Datasets** | **Name of the methods** | | | | |
|---|---|---|---|---|---|---|
| | | ***SAS-DPSO*** | ***CDNNCS*** | ***J-SSO*** | ***GC*** | ***CGBFO-GC (proposed)*** |
| KPA | Wholesale customer data | 0.9999 | 1 | 1 | 1 | 0.99799 |
| | Heart disease data | 0.99698 | 0.9983 | 0.99972 | 0.99979 | 0.99498 |
| | Air quality datasets | 0.9998 | 0.9992 | 0.9998 | 0.9999 | 0.9997 |
| CPA | Whole sale customer data | 1 | 1 | 1 | 1 | 0.998 |
| | Heart disease data | 0.9993 | 0.99986 | 0.9994 | 0.9993 | 0.99786 |
| | Air quality datasets | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.99799 |

CDNNCS, J-SSO, GC, and proposed CGBFO-GC methods. However, the proposed CGBFO-GC method demonstrates 0.998%, 0.9978% and 0.99799% correlation values than previous studies. To secure the cloud data, the proposed CGBFO-GC method is more effective against attacks while contrasted to the other methods.

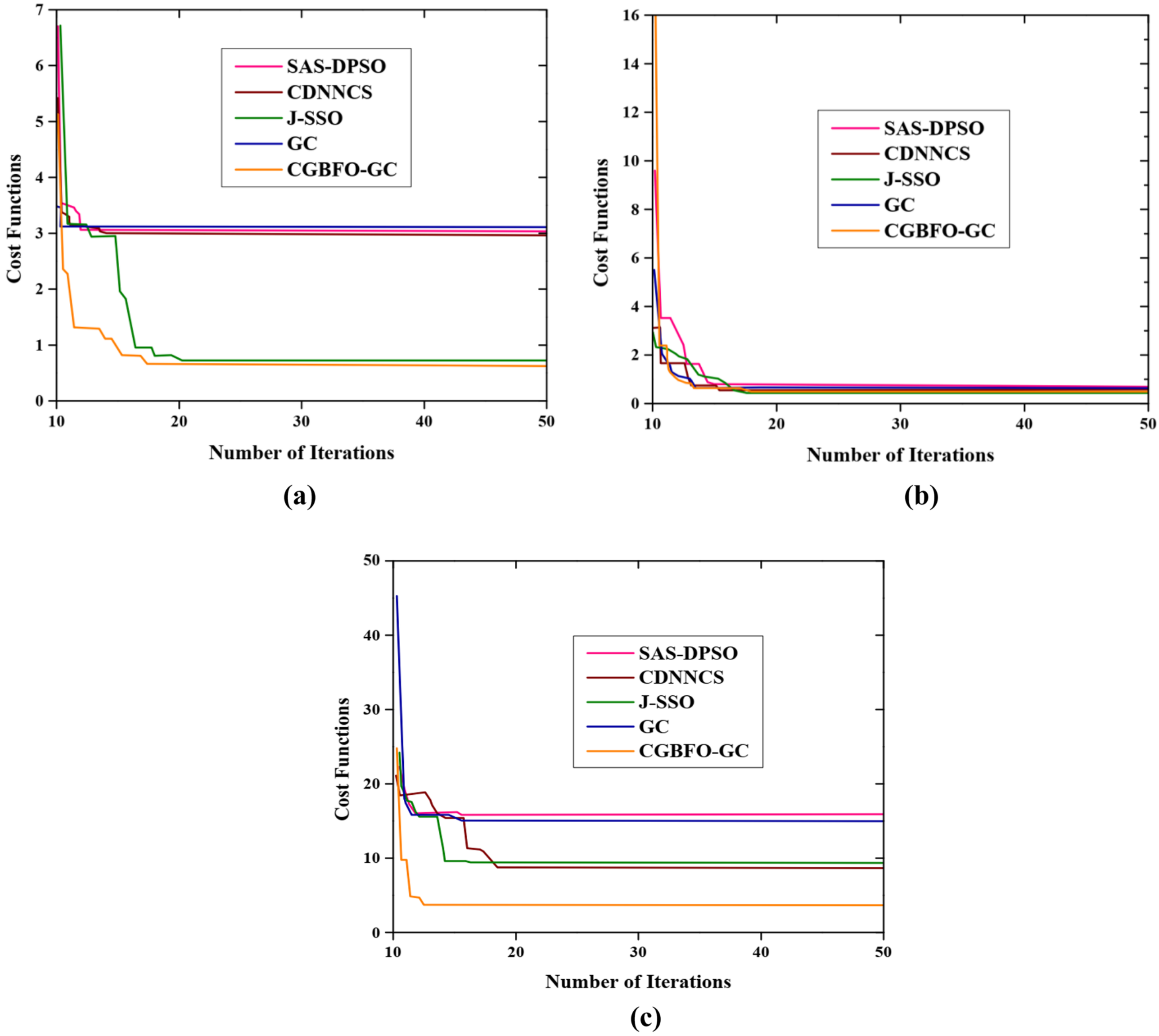


**Fig. 7** Illustration of proposed CGBFO-GC based cloud data security, (**a**) Dataset-1, (**b**) Dataset-2, and (**c**) Dataset-3

**Table 3** Proposed cloud data security with the evaluation of key sensitivity based on a different dataset

| **Name of the methods** | **Name of the datasets** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Dataset-1 (%)** | | | | **Dataset-2 (%)** | | | | **Dataset-3 (%)** | | | |
| | **15** | **30** | **45** | **60** | **15** | **30** | **45** | **60** | **15** | **30** | **45** | **60** |
| ***SAS-DPSO*** | 0.999 | 1 | 0.996 | 0.990 | 0.997 | 0.995 | 0.994 | 0.63 | 0997 | 0.999 | 0.984 | 0.993 |
| ***CDNNCS*** | 0.998 | 1 | 0.999 | 0.999 | 0.999 | 0.899 | 0.924 | -0.2 | 0.999 | 0.997 | 0.4002 | 0.1957 |
| ***J-SSO*** | 1 | 0.99 | 0.99 | 1 | 0.906 | 0.985 | 0.619 | 0.834 | 0.7314 | 0.3182 | 0.997 | 0.989 |
| ***GC*** | 0.999 | 1 | 0.999 | 0.994 | 0.993 | 0.944 | -0.438 | 0.860 | 0.993 | 0.998 | 0.997 | 0.995 |
| ***CGBFO-GC (proposed)*** | 0.997 | 0.999 | 0.996 | 0.993 | -0.006 | 0.9905 | -0.530 | -0.395 | 0.954 | 0.993 | -0.245 | -0.434 |

## 5.5 Literature bio-inspired models with its comparative analysis

The performance of the proposed CGBFO-GC based cloud data security model is demonstrated in Fig. 7a–c by comparing it to conventional techniques in terms of datasets 1, 2, and 3. In this experiment, the graph is plotted between a number of iterations and cost functions. The proposed method is 97% better than SAS-DPSO, CDNNCS, J-SSO, and GC models when the iteration is 40 for wholesale customer data. While comparing with the related works, the proposed method demonstrates efficient cloud data security.

## 5.6 Evaluation of key sensitivity

Table 3 describes the novel cloud data security with key sensitivity evaluation based on various datasets. To evaluate the key sensitivity, we execute 15%, 30%, 45% and 60% variations. The key had minimal variation, and the analysis provided the relationship between the original and restored data. The table below shows the key sensitivity results for datasets 1, 2, and 3. However, the proposed CGBFO-GC technique accomplishes superior performances in terms of key sensitivity evaluation.

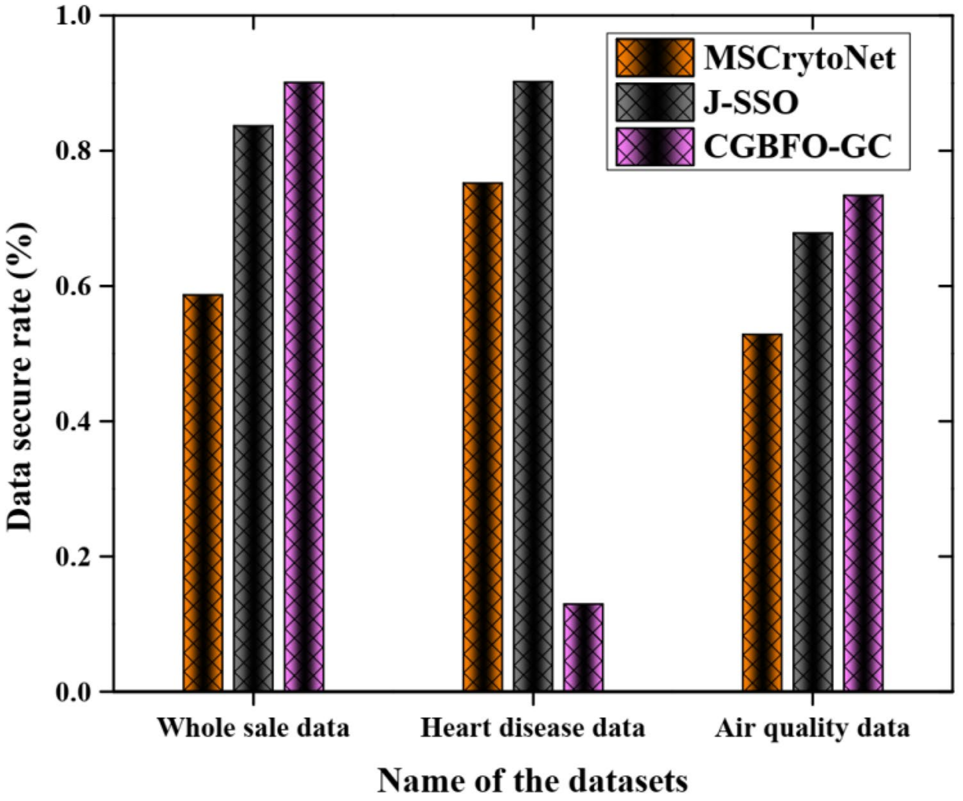


**Fig. 8** State-of-art comparison of the different data set with a secure data rate

## 5.7 State-of-art comparison

Figure 8 illustrates the state-of-the-art comparison of different datasets with secure data rates. The comparative analysis is performed using MSCryptoNet, J-SSO, and the proposed CGBFO-GC technique. For wholesale data, we have obtained 0.587%, 0.837%, and 0.901% secure data rates for MSCryptoNet, J-SSO, and CGBFO-GC methods. Further, the MSCryptoNet, J-SSO and CGBFO-GC method demonstrate 0.752%, 0.902% and 0.1293% for heart disease data. Similarly, we have obtained 0.528%, 0.678%, and 0.7340% for MSCryptoNet, J-SSO, and CGBFO-GC methods. The proposed CGBFO-GC method, on the other hand, outperforms existing methods in terms of cloud data security.

## 5.8 Computational time analysis

Figure 9 illustrates the state-of-the-art comparison in terms of computational time analysis. This graph depicts the relationship between different methodologies and computation time, with computation time measured in seconds (sec). The SAS-DPSO, CDNNCS, J-SSO, GC, and proposed CGBFO-GC demonstrate 134.72 s, 142.38 s, 129.78 s, 159.11 s, and 103.21 s computational time. When compared to all the existing methods, the proposed CGBFO-GC algorithm provides lower computational time with better speed.

## 5.9 Big 'O' notation for proposed CGBFO-GC time complexity analysis

The execution time complexity is analyzed using the most common metric called Big 'O'. In this study, $O(Max_{itr} * SP \wedge 2)$ is the time complexity of the proposed CGBFO-GC algorithm. Where the maximum iteration is $Max_{itr}$ and population size is $SP$. The proposed CGBFO-GC algorithm with its space complexity is $O(Max_{itr} * SP^2)$.

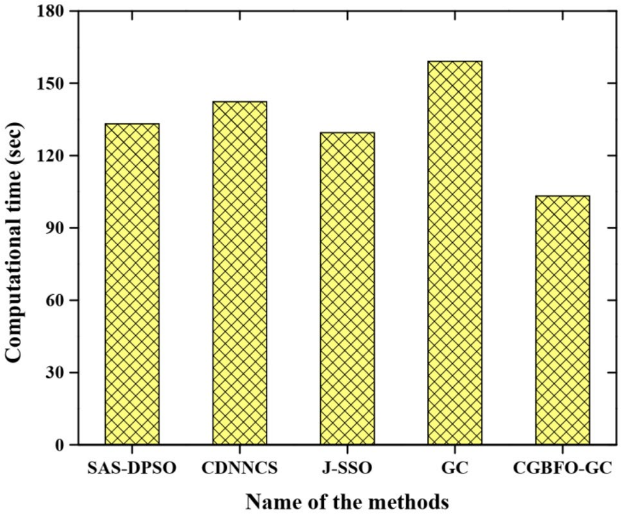


**Fig. 9** State-of-art comparison of computational time analysis

## 6 Conclusion

This paper presented a CGBFO- GC algorithm-based privacy preservation model for the cloud. In the proposed model, the major steps are information sanitation and restoration with optimal key generation. The multi-objective parametric function such as hiding ratio, data preservation ratio, and modification degree derives the CGBFO-GC algorithm optimizes the optimal key. While dealing with varied keys and different attacks, the proposed method contains a fast convergence rate with the ability in solving multi-objective privacy preservation issues. The proposed CGBFO- GC has a lower computation time than existing methods such as SAS-DPSO, CDNNCS, J-SSO, and GC. In terms of key sensitivity evaluation, CPA and KPA attacks analysis, and convergence analysis, the proposed method achieves optimal and superior outcomes when compared to existing algorithms. The proposed model has a few shortcomings, such as slight data loss and a somewhat expensive security model. As a result, we propose to provide an efficient deep learning model to overcome the aforementioned drawbacks with an optimization technique and a cryptography model in future work.

**Availability of data and material** Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## Declarations

**Conflict of interest** The authors declare that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

## References

1. Sabin Begum R, Sugumar R (2019) Novel entropy-based approach for cost-effective privacy preservation of intermediate datasets in cloud. Clust Comput 22(4):9581–9588
2. Kanwal T, Anjum A, Khan A (2020) Privacy preservation in e-health cloud: taxonomy, privacy requirements, feasibility analysis, and opportunities. Clust Comput 1–25
3. Kaaniche N, Laurent M (2017) Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms. Comput Commun 111:120–141
4. Ma Z, Ma J, Miao Y, Liu X, Choo KKR, Yang R, Wang X (2020) Lightweight privacy-preserving medical diagnosis in edge computing. IEEE Trans Serv Comput
5. Fang C, Guo Y, Wang N, Ju A (2020) Highly efficient federated learning with strong privacy preservation in cloud computing. Comput Secur 96:101889
6. Berlato S, Carbone R, Lee AJ, Ranise S (2020, October). Exploring Architectures for Cryptographic Access Control Enforcement

in the Cloud for Fun and Optimization. In Proceedings of the 15th ACM Asia Conference on Computer and Communications Security (pp. 208–221)

7. Shankar K (2021) Improving the Security and Authentication of the Cloud with IoT using Hybrid Optimization Based Quantum Hash Function. Journal of Intelligent Systems and Internet of Things 1(2):61–71
8. 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
9. Sundararaj V, Anoop V, Dixit P, Arjaria A, Chourasia U, Bhambri P, Rejeesh MR, Sundararaj R (2020) CCGPA-MPPT: Cauchy preferential crossover-based global pollination algorithm for MPPT in photovoltaic system. Prog Photovolt Res Appl 28(11):1128–1145
10. Vinu S (2019) Optimal task assignment in mobile cloud computing by queue based ant-bee algorithm. Wirel Pers Commun 104(1):173–197
11. 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
12. Rejeesh MR (2019) Interest point based face recognition using adaptive neuro fuzzy inference system. Multimed Tools Appl 78(16):22691–22710
13. Gowthul Alam MM, Baulkani S (2019) Geometric structure information based multi-objective function to increase fuzzy clustering performance with artificial and real-life data. Soft Comput 23(4):1079–1098
14. Gowthul Alam MM, Baulkani S (2019) Local and global characteristics-based kernel hybridization to increase optimal support vector machine performance for stock market prediction. Knowl Inf Syst 60(2):971–1000
15. Hassan BA (2020) CSCF: a chaotic sine cosine firefly algorithm for practical application problems. Neural Comput Appl 1–20
16. Hassan BA, Rashid TA (2021) A multidisciplinary ensemble algorithm for clustering heterogeneous datasets. Neural Comput Applic 1–24
17. Haseena KS, Anees S, Madheswari N (2014) Power optimization using EPAR protocol in MANET. IJISET-International Journal of Innovative Science, Engineering & Technology 1(6)
18. Jose J, Gautam N, Tiwari M, Tiwari T, Suresh A, Sundararaj V, Rejeesh MR (2021) An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion. Biomed Signal Process Control 66:102480
19. Anand K, Vijayaraj A, Vijay Anand M (2022) Privacy preserving framework using Gaussian mutation based firebug optimization in cloud computing. J Supercomput 78:9414–9437. https://doi.org/10.1007/s11227-021-04173-w
20. Ahamad D, Hameed SA, Akhtar M (2020) A multi-objective privacy preservation model for cloud security using hybrid Jaya-based shark smell optimization. Journal of King Saud University-Computer and Information Sciences
21. Ahmad W, Wang S, Ullah A, Mahmood Z (2018) Reputation-aware trust and privacy-preservation for mobile cloud computing. IEEE Access 6:46363–46381
22. Li J, Kuang X, Lin S, Ma X, Tang Y (2020) Privacy preservation for machine learning training and classification based on homomorphic encryption schemes. Inf Sci 526:166–179
23. Abirami P, Vijay Bhanu S (2020) Enhancing cloud security using crypto-deep neural network for privacy preservation in trusted environment. Soft Comput 24(24):18927–18936
24. Park JY, Huh EN (2020) A Cost-Optimization Scheme Using Security Vulnerability Measurement for Efficient Security Enhancement. J Inf Process Syst 16.1
25. Meng S, Huang W, Yin X, Khosravi MR, Li Q, Wan S, Qi L (2020) Security-aware dynamic scheduling for real-time optimization in cloud-based industrial applications. IEEE Trans Ind Inf
26. Sugumaran M, Bala Murugan B, Kamalraj D (2014) An architecture for data security in cloud computing. In 2014 World Congress on Computing and Communication Technologies, pp. 252–255. IEEE
27. Gupta I, Singh N, Singh AK (2019) Layer-based privacy and security architecture for cloud data sharing. J Commun Softw Syst 15(2):173–185
28. Cardoso Margarida GMS (2013) Logical discriminant models â€“ Chapter 8 in Quantitative Modeling in Marketing and Management Edited by Luiz Moutinho and Kun-Huang Huarng. World Scientific. p. 223–253. ISBN 978–9814407717
29. Wl odzisl and Rafal Adamczak and Krzysztof Grabczewski and Grzegorz Zal. A hybrid method for extraction of logical rules from data. Department of Computer Methods, Nicholas Copernicus University
30. De Vito S, Massera E, Piga M, Martinotto L, Di Francia G (2008) On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. Sensors Actuators B Chem 129(2):750–757. ISSN 0925–4005
31. Rajasekar N, Kumar NK, Venugopalan R (2013) Bacterial foraging algorithm based solar PV parameter estimation. Sol Energy 97:255–265
32. Xu Y, Chen H, Heidari AA, Luo J, Zhang Q, Zhao X, Li C (2019) An efficient chaotic mutative moth-flame-inspired optimizer for global optimization tasks. Expert Syst Appl 129:135–155
33. Yue X, Zhang H (2020) Modified hybrid bat algorithm with genetic crossover operation and smart inertia weight for multilevel image segmentation. Appl Soft Comput 90:106157
34. Manogaran G, Thota C, Kumar MV (2016) MetaCloudDataStorage architecture for big data security in cloud computing. Procedia Comput Sci 87:128–133
35. Talib AM, Atan R, Abdullah R, Murad MAA (2010) Security framework of cloud data storage based on multi agent system architecture: Semantic literature review. Comput Inform Sci 3(4):175
36. Ramasamy S, Gnanamurthy RK (2020) Cluster Based Multi Layer User Authentication Data Center Storage Architecture for Big Data Security in Cloud Computing. J Internet Technol 21(1):159–171
37. Sharma S, Mishra A, Singhai D (2020) Secure cloud storage architecture for digital medical record in cloud environment using blockchain
38. Thakare VR, Singh KJ (2020) Cloud Security Architecture Based on Fully Homomorphic Encryption. In Architecture and Security Issues in Fog Computing Applications, pp. 83–89. IGI Global
39. Ghani A, Badshah A, Jan S, Alshdadi AA, Daud A (2020) Cloud storage architecture: research challenges and opportunities. arXiv preprint arXiv:2004.06809
40. Kumar YK, Shafi RM (2020) An efficient and secure data storage in cloud computing using modified RSA public key cryptosystem. Int J Electr Comput Eng 10(1):530
41. Malik A, Aggarwal M, Sharma B, Singh A, Singh KK (2020) Optimal Elliptic Curve Cryptography-Based Effective Approach for Secure Data Storage in Clouds. Int J Knowl Syst Sci (IJKSS) 11(4):65–81
42. Murugesan A, Saminathan B, Al-Turjman F, Kumar RL (2020) Analysis on homomorphic technique for data security in fog computing. Trans Emerg Telecommun Technol e3990

**Dr. K. Anand** obtained his Master's degree in Computer Science and Engineering and Ph.D degree specialized in Network Security from Anna University in the year 2019. He is currently working as Associate Professor in the Department of Computer Science and Engineering, Chennai Institute of Technology, Chennai, Tamil Nadu, India. His specialization is Network Security and Information Retrieval, Web Mining.

**Dr. A. Vijayaraj** is an Associate Professor; Department of Information Technology, Vignan's Foundation for Science, Technology & Research (Deemed to be University) Vadlamudi, Guntur, Andhra Pradesh from August 2020. He obtained his Bachelor's degree from Bharathidhasan University, in 1997 and his Master of Engineering in Computer Science and Engineering from Sathyabama University in 2005. He obtained his Ph.D from Anna University, India. His area of specialization is networks and communication, operating systems, mobile computing, Information Retrieval, Knowledge Management. He has 20 years of teaching experience from various Engineering Colleges during tenure he was Awarded Best Teacher Award thrice. He is a Member of CSI, ISTE, IAENG, ICST, UACEE, IASTER and CSTA. He organized number of Workshops, Faculty development programs, Seminars, National and International conferences. He has Published 30 papers in various reputed International journals and 22 Papers in International and National Level conferences.

**M. Vijay Anand** obtained his Bachelor's degree in Computer Science and Engineering from Madras University. Then he obtained his Master's degree in Computer Science and Engineering from Sathyabama Institute of Science and Technology. He obtained his Ph.D from Anna University, India. Currently, he is a Professor in Computer Science and Engineering at Saveetha Engineering College India. His specializations include Networking, Manet, Network Security, Wireless Sensor Network.