Hybrid optimization-based privacy preservation of database publishing in cloud environment

Article *in* Concurrency and Computation Practice and Experience · February 2022

DOI: 10.1002/cpe.6844

J Kingsleen Solomon Doss
Vels University

Kamalakkannan Somasundaram
Vels University

Received: 12 August 2021 | Revised: 10 November 2021 | Accepted: 7 December 2021
DOI: 10.1002/cpe.6844

WILEY

RESEARCH ARTICLE

Hybrid optimization-based privacy preservation of database publishing in cloud environment

Kingsleen Solomon Doss[1] | Somasundaram Kamalakkannan[2]

[1]Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Pallavaram Chennai, India

[2]Department of Information Technology, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, India

Correspondence
Kingsleen Solomon Doss, Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Pallavaram, Chennai, India.
Email: kingsleensolomon@gmail.com

Abstract

This work intends to introduce an optimization-based privacy preservation model via selecting the optimal key matrix. Here, privacy preservation is carried out under two processes, namely, "data sanitization and restoration." In fact, data sanitization is the data preservation method, where the data (message) are preserved using the optimal key. Similarly, data restoration is the inverse procedure of sanitization. Here, the key matrix is optimally chosen using a novel hybrid algorithm. For optimization purpose, this work deploys a hybrid optimization approach known as random-based grey dragon algorithm (R-GDA) that involves the concepts of both "dragonfly algorithm (DA) and grey wolf optimization (GWO) algorithm." The novelty of the work is introduced in hybrid optimization approach R-GDA. Eventually, the supremacy of the adopted method is validated over other existing approaches in terms of various measures such as privacy, utility, and so on. The privacy preservation in the cloud is achievable in the field of education, banking sector, military, and the research community.

KEYWORDS
cloud computing, key matrix, privacy, R-GDA algorithm, sanitization

1 | INTRODUCTION

Privacy preservation in the cloud paradigm is emerging as a most important research problem in present days. Cloud environments handle the information from vast users through remote platforms in a simultaneous manner.[1–3] The cloud server enables users for utilizing the data services, and users could share the information with clientele. In addition, the cloud platform permits users to store, edit and recover a huge quantity of data. Owing to extensive appliances offered by cloud paradigm, private firms, and official firms access the cloud platforms for varied applications.[4,5]

In fact, the information given by users is sensitive, and therefore, the data should be conserved before being declared to public. The cloud platforms provide a shared data pool to certified users, and the service offered to the users are private/public.[6,7] Moreover, it is essential for cloud servers to recognize authentic user, and consent to data access. High perceptive appliances like research and military community, necessitate more confidentiality, and moreover, a slighter modification in the original information during the data recovery might corrupt the performances severally.[8,9] The challenges issues include handling of incomplete or delayed information, infinite data length, handling of the steady stream of information as well as privacy preservation of the data in the process of extraction from the stream of data. In the cloud environment the presence of a large number of users in the data pool, and hence, it is extremely difficult to maintain the privacy of every database.[2]

Privacy preservation techniques are dependent on data perturbation, data encryption, and so on.[10–12] Preserving the confidentiality has attained more recognition as it effectively hides the confidential data, and it offers the maximal utility throughout the retrieval.[13,14] Encryption oriented approaches, namely, ECC and homomorphic encryption, also contributed to the privacy conservation, and they modified the database using encryption oriented approaches.[15–17]

https://doi.org/10.1002/cpe.6844

The major contribution of research is described here:

- Introduces a novel privacy preservation model, where hybrid optimization is deployed for producing the optimal key matrix.
- The input database from the user is integrated with the secret key obtained by the proposed R-GDA for generating the retrievable database. One of the major aims of the proposed scheme is to preserve the input data.
- The R-GDA uses the fitness function for deriving the secret key, such that the privacy and the utility of the data are maintained as high as possible. The secret key generated by the optimization can be used at data retrieval phase by the client, to get access to the data.
- Proposes a new random-based grey dragon algorithm, which is the hybridized version of existing GWO and DA models.

The rest of this work is arranged as: Section 2 analyzed the review on related topic. Section 3 portrays the constructing retrievable perturbation database and Section 4 depicts the proposed random-based grey dragon algorithm for optimal key matrix. Reconstructing the original data from perturbed data is described in Section 5. The results are exhibited in Section 6 and Section 7 portrays the conclusion.

2 | LITERATURE REVIEW

2.1 | Related Works

In 2019, Chunhui et al.[18] discussed the privacy risks and presented a differential privacy approach for publishing government data depending on fog computing. Accordingly, a data publishing model was developed by means of MaxDiff histogram for realizing the user privacy preserving function depending on fog environment. Finally, the simulated experimentations have revealed that the presented method offered enhanced performance in terms of query sensitivity.

In 2019, Revathi et al.[19] proposed BS-WOA for recognizing the confidential key. The database from user was updated with optimum confidential key for preserving the utility and privacy of data. The adopted BS-WOA was modeled via the hybridization of BSO and WOA. Finally, by carrying out a comprehensive examination, the enhancement of the developed technique was verified regarding privacy.

In 2016, Kan et al.[20] have proposed a privacy-preserving AKPS method for cloud environment. Particularly, the attribute-oriented encryption was employed for encrypting the published data, by which the publishers controlled the data access by themselves. In addition, a novel searchable encryption model was presented, by which the subscribers' can selectively receive the interested data. Finally, security analysis has revealed a strong security assurance with a higher effectiveness of the developed model.

In 2019, Abdul[21] has proposed a novel anonymization system of data confidentiality for EHR that differed from conventional schemes by its capability to avoid adversaries. The developed technique transformed data into predetermined intervals and subsequently, the original values were replaced with averages. Therefore, the adopted model offered enhanced data utility in preserving data publishing. The simulated outcomes have shown the efficiency of presented model with respect to enhanced privacy.

In 2020, Shani et al.[22] have proposed an effectual "structured data sharing framework" for revocation of users in cloud system. The adopted technique depends on the five parameters, namely, "data owner, cloud storage, central authority, and cryptographic server, and data users." The user then submits user list that included the capability of producing an ACL. At the end, the supremacy of proposed approach was validated over the extant schemes.

In 2020, Sana et al.[23] have proposed a privacy preserving approach termed as Ins-PAbAC, which securely shared the outsourced data content through public cloud servers. The developed scheme presented numerous benefits. Initially, it provided encrypted access control features and, moreover, it preserved privacy of users based on authentication mechanism. In addition, attribute-based signature was introduced, which revealed the identities of anonymously-authenticated users if required. Furthermore, the adopted model was resistant to malicious adversaries in the cloud servers.

In 2020, Sudhakar et al.[24] have proposed a proficient "index oriented quasi-identifier approach" that ensured privacy and achieved higher data utility over distributed and incremental data sets. The updated FCM approach was exploited for forming the similarity-based clusters. Finally, the analysis outcomes have illustrated that the presented technique was more effectual for conserving privacy when compared to existing ones.

In 2020, Bibal et al.[25] have employed k-anonymization model for upgrading the privacy strategies in cloud storages. Along with this, the GGWO model was deployed that decided the data to be published depending on the data conserved for confidentiality purposes. In the end, evaluation was carried out that proved the improvement of the adopted approach over existing models in terms of information loss, utility and privacy.

In 2013, John et al.[26] have proposed a privacy preservation model with the Naive Bayesian classification algorithm. The cryptographic technique implemented is RSA. Finally, the analysis outcomes are proved out interms of accuracy.

TABLE 1 Review on traditional privacy preservation techniques in cloud

Authors	Adopted model	Features	Challenges
Chunhui et al.[18]	MaxDiff histogram	• Minimal average error • Reduced query sensitivity	• Data sharing is not carried out in real time.
Revathi et al.[19]	BS-WOA	• High privacy • High utility value	• No consideration on real time execution
Kan et al.[20]	AKPS scheme	• Practically efficient • Minimal time consumption	• Need to support keyword queries
Abdul[21]	Anonymization scheme	• Conserves better utility • Offers better privacy	• Secure provenance remains as an issue
Shani et al.[22]	Hash algorithm	• Minimal execution time • Reduced memory usage	• Have to cope up with insider threats
Sana et al.[23]	Ins-PAbAC architecture	• Minimal time consumption • Minimal communication cost	• Needs exploration on decryption outsourcing models
Sudhakar et al.[24]	FCM Algorithm	• Highly efficient • Offers reduced processing time	• The cluster similarity was formulated only for arithmetical data
Bibal et al.[25]	GGWO algorithm	• Minimal information losses • High utility	• Needs to exploit varied datasets for publishing phase
John et al.[26]	Naive Bayesian algorithm	• The computational complexity is less • Simple method	• The privacy preservation of data is not as much as safe.

2.2 | Review

Table 1 explains the characteristics and challenges in traditional methods regarding the privacy preservation of data in cloud environment. More research works are exploited regarding this concept and the methodologies related with their works with their pros and cons are explained as follows: MaxDiff histogram[18] has reduced query sensitivity with minimal average error; however, data sharing is not carried out in real time. BS-WOA[19] improves the privacy and satisfies all requirements of utility value. However, need to explore more on real time execution. AKPS scheme[20] is practically efficient and poses minimal time consumption. Further it needs to support keyword queries. Anonymization scheme[21] has resulted in better utility and improved privacy, yet secure provenance remains as an issue. Hash algorithm[22] ensures reduced memory usage with minimal execution time; however, it should cope up with insider threats. Ins-PAbAC architecture is used in Reference 23 offers minimal time consumption with minimal communication cost. However, it needs exploration on decryption outsourcing models. FCM algorithm adopted in Reference 24 is very much efficient and offers reduced processing time. Nevertheless, cluster similarity was formulated only for arithmetical data. GGWO algorithm used in Reference 25 offers minimal information losses with high utility, but it needs to exploit varied datasets for publishing phase.

2.3 | Problem statement

The difficulties involved in the construction of the privacy preserved database through the optimization-based schemes are given. The cloud users can manage a large number of users at a same time. The most important factor is the amount of trust level among the users.[20] For preserving the data, encryption and decryption are used. During decryption, the original quality of the data is degraded.[10,23] Hence, the privacy preservation model needs to ensure both the privacy and the utility of the data structure. Since modifying the input data should not alter the numerical characteristics. The privacy rate should depend on the trust level of the user.[27]

3 | IMPLEMENTED PRIVACY PRESERVATION MODEL IN CLOUD ENVIRONMENT

Figure 1 shows the diagrammatic representation of the developed privacy preservation framework using new R-GDA algorithm. Initially, the original database is combined with the optimal key matrix provided by the R-GDA based on Tracy Singh Product. The matrix thus attained has a bigger size, and therefore the matrix is minimized to the size identical to input data. Subsequently, a secret key is selected from the reduced matrix, whose coefficients hold a most important part in constructing the recoverable perturbation database. Furthermore, the service provider could access the original data by using confidential key.

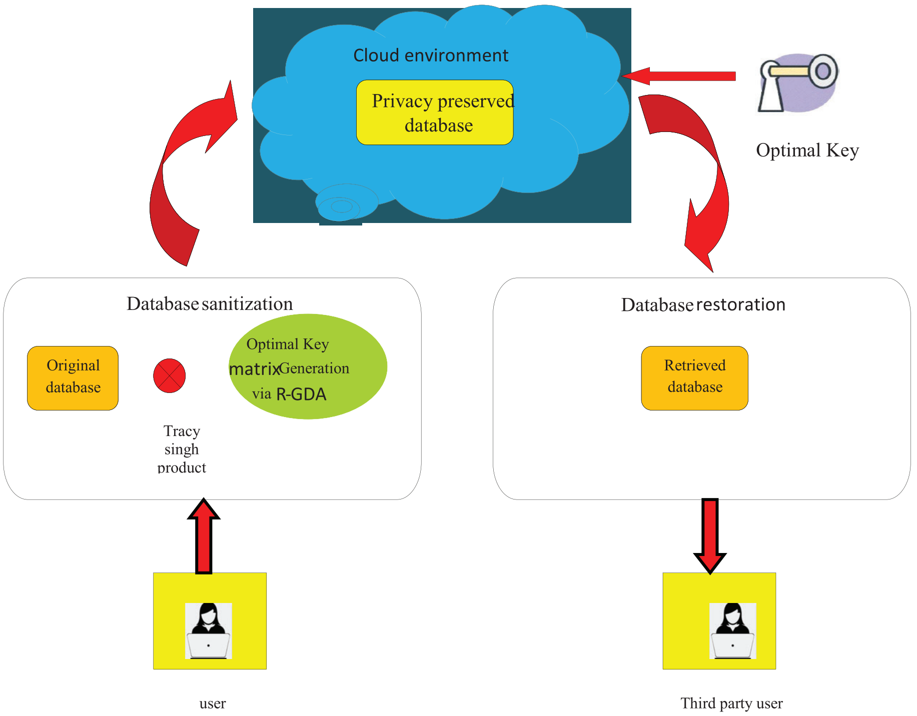


FIGURE 1 Block diagram of developed privacy preservation model

3.1 | Constructing retrievable database

Assume the database offered by data owner as A and data matrix size gets varied as $P * Q$. The demonstration of input data matrix is shown in Equation (1), in which a_{ij} points out the data coefficients and j and i alter from Q to P in that order.

$$A_{P \times Q} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} \\ a_{21} & a_{22} & \dots & a_{2j} \\ \vdots & & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} \end{bmatrix} \tag{1}$$

Subsequently, the matrix given as input is factorized by the TSP[28,29] with optimum key matrix produced by the developed R-GDA model. The depiction of the TSP is signified in Equation (2), wherein $B_{R \times S}$ points out the optimal key matrix attained from R-GDA that is depicted in Equation (3).

$$C_{PR \times QS} = A_{P \times Q} \circ B_{R \times S} \tag{2}$$

$$B_{R \times S} = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1l} \\ b_{21} & b_{22} & \dots & b_{2l} \\ \vdots & & & \vdots \\ b_{k1} & b_{k2} & \dots & b_{kl} \end{bmatrix} \tag{3}$$

In Equation (3), b_{kl} points out the optimal key matrix coefficients. Equation (4) shows the model for Tracy Singh product among the input data matrix and optimal key matrix.

$$M_{XA \times NB} = \left(A_{ij} \circ B\right)_{ij} = \left(\left(A_{ij} \otimes B_{kl}\right)_{kl}\right)_{ij} \tag{4}$$

Furthermore, the key coefficients are multiplied with partial data matrix of i/p data as specified in Equation (5).

$$M_{XA\times NB} = \begin{bmatrix} A_{11}\circ B & A_{12}\circ B \\ A_{21}\circ B & A_{22}\circ B \end{bmatrix} \tag{5}$$

In Equation (5), A_{11}, A_{12}, A_{21}, and A_{22} belong to i/p data matrix. Subsequently, the size of Tracy Singh product is minimized for constructing a condensed matrix as specified in Equation (6).

$$Z_{P\times Q} = \begin{bmatrix} Z_1 & Z_2 \\ Z_3 & Z_4 \end{bmatrix} \tag{6}$$

In Equation (7), Z_1, Z_2, Z_3, and Z_4 point out the data matrix components. Finally, the secret key $D_{1\times1} = a_{11}$ computed from the optimal key matrix, that is, initial optimal key matrix element, carries out EX-OR function with condensed matrix to recognize the perturbed database $A*$ as shown in Equation (7), where $Z_{P\times Q}$ points out the condensed matrix.

$$A^{*}_{P\times Q} = Z_{P\times Q} \oplus D_{1\times1} \tag{7}$$

The example for constructing a condensed matrix using the Tracy Singh product is depicted below:

Consider $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ as original matrix and let $\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix}$ be the key matrix. On multiplying the original matrix with key matrix using Tracy Singh product, a 4×4 matrix is formed as shown in Equation (8).

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix} = \begin{bmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{bmatrix} \tag{8}$$

This 4×4 matrix is then sub-divided into four 2×2 matrix as $\begin{bmatrix} 0 & 5 \\ 6 & 7 \end{bmatrix}\begin{bmatrix} 0 & 10 \\ 12 & 14 \end{bmatrix}\begin{bmatrix} 0 & 15 \\ 18 & 21 \end{bmatrix}\begin{bmatrix} 0 & 20 \\ 24 & 28 \end{bmatrix}$. As the initial key matrix element is zero, the second key matrix element is chosen from 4 subdivided 2×2 matrices. Thereby, the condensed matrix will be formed as shown in Equation (9).

$$\begin{bmatrix} 5 & 10 \\ 15 & 20 \end{bmatrix} \tag{9}$$

4 | PROPOSED RANDOM-BASED GREY DRAGON ALGORITHM FOR OPTIMAL KEY MATRIX SELECTION

4.1 | Solution encoding and objective function

The presented work concerns on optimal selection of key matrix using the R-GDA algorithm. The solutions provided for encoding is shown in Figure 2, where *H* points out the key matrix and *nu* refers to the total number of key matrices. The objective function of the developed work is shown in Equation (10), in which Pr points out the privacy. The privacy Pr is computed as shown in Equation (11).

$$Ob = Min\left[\frac{1}{\mathrm{Pr}}\right] \tag{10}$$

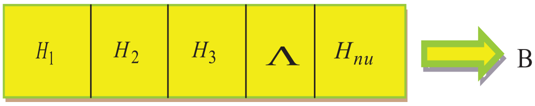


FIGURE 2 Solution encoding

$$\Pr = \frac{1}{P \times Q} \sum_{i=1}^{P} \sum_{j=1}^{Q} \frac{\left(A_{ij} - A_{ij}^{*}\right)}{Max\left(A_{ij}, A_{ij}^{*}\right)} \tag{11}$$

4.2 | Proposed R-GDA algorithm

Although GWO[27] offers enhanced qualities like flexibility and stability; it endures from slower convergence. Hence, the concept of DA[30] is mingled with it to introduce a new algorithm. Hybrid optimization algorithms have been reported to be promising for certain search problems.[17,31–33] The procedure of the proposed R-GDA model is as follows. As per the developed model, if the random integer, r is greater than 0.5, the update takes place based on GWO algorithm. The wolves α, β, and γ are the most important wolves, which concern on hunting process. Among these, α is the leader that takes decisions on hunting procedure, sleeping place, time for awakening, and so on; while β and γ take the second and third levels that aid α in making decisions. Accordingly, the last level of wolves is ζ that focuses on eating. The encircling feature of wolves is formulated as in Equation (13), where U signify coefficient vectors, B_p points out position vector of prey, B points out position vector of wolves, it points out present iteration, and Y refers to the distance. In Equation (12), F denotes the random vector that lies between 0 and 1 and it is computed as per Equation (15). The constraint $\widehat{b}$ in Equation (14) lies between 2 and 0. Here, ra_1 and ra_2 point out the random vectors between [0, 1] and $it_{\max}$ points out the maximal iteration.

$$Y = \left|F.B_p(it) - B(it)\right| \tag{12}$$

$$B(it+1) = B_p(it) - U.Y \tag{13}$$

$$U = 2\widehat{b}.ra_1 - \widehat{b} \tag{14}$$

$$F = 2ra_2 \tag{15}$$

The numerical formulation for relating the chasing nature of wolf is specified in Equations (16) to (21). Accordingly, the final update takes place as shown in Equation (22).

$$Y_{\alpha} = |F_1.B_{\alpha} - B| \tag{16}$$

$$Y_{\beta} = \left|F_1.B_{\beta} - B\right| \tag{17}$$

$$Y_{\gamma} = \left|F_1.B_{\gamma} - B\right| \tag{18}$$

$$B_1 = B_{\alpha} - U_1.(Y_{\alpha}) \tag{19}$$

$$B_2 = B_{\beta} - U_2.\left(Y_{\beta}\right) \tag{20}$$

$$B_3 = B_{\gamma} - U_3.\left(Y_{\gamma}\right) \tag{21}$$

$$B(it+1) = \frac{B_1 + B_2 + B_3}{3} \tag{22}$$

On the other hand, if the random integer r is less than 0.5, the position gets updated based on the levy update of DA as per Equation (23), in which z signifies dimension.

$$B(it+1) = B(it) + Levy(z) \times B(it) \tag{23}$$

The Levy flight is evaluated by Equation (24), wherein η specifies a constant variable and y_1 and y_2 are the arbitrary integers. δ is computed as shown in Equation (25), where $\Gamma(x) = (x-1)$ and η points out a constant.

$$Levy(x) = 0.01 \times \frac{y_1 \times \delta}{|y_2|^{\frac{1}{\eta}}} \tag{24}$$

$$\delta = \left(\frac{\Gamma(1+\eta) \times \sin\left(\frac{\pi\eta}{2}\right)}{\Gamma\frac{(1+\eta)}{2} \times \eta \times 2^{\left(\frac{\eta-1}{2}\right)}} \right)^{\frac{1}{\eta}} \tag{25}$$

The pseudo code of developed R-GDA scheme is highlighted in Algorithm 1.

Algorithm 1. Developed R-GDA Model

```
Initialization
Compute all search agent's fitness
Assign $B_\alpha$, $B_\delta$ and $B_\delta$ as 1st, 2nd and 3rd best agents
While $(it < it_{\max})$
        For every wolf
        If $r < 0.5$
            Update position based on GWO as per Equation (22)
        Else
            Update position based on DA as per Equation (23)
        End for
        Update $\hat{b}$, $U$ and $F$
        Compute all search agent's fitness
        Update $B_\alpha$, $B_\delta$ and $B_\delta$
        $it = it + 1$
End while
Return $B_\alpha$
```

In the metaheuristic approach, the hybridization concept means merging the algorithms in order to provide a new powerful algorithm based on the features of the merged ones. In the adopted model, GWO is merged with DA. The major drawback of the GWO is that it endures lower convergence and fall on local optima. To overcome the certain drawback, the GWO is hybridized with DA. The DA will provide better convergence and it will avoid local minima that may cause premature convergence

5 | RECONSTRUCTING THE ORIGINAL DATA FROM PERTURBED DATA

During retrieval, the perturbed data in cloud are carried out using EX-OR function with the confidential key that results in condensed matrix $Z^*_{P\times Q}$. Furthermore, dividing the condensed matrix $Z^*_{P\times Q}$ with confidential key D helps in finding the appropriate original database at receiving side. In Equation (27), $A^*_{P\times Q}$ points out the retrievable perturbation matrix.

$$Z^*_{P\times Q} = A^*_{P\times Q} \oplus D_{1\times 1} \tag{26}$$

$$A^*_{P\times Q} = \frac{Z^*_{P\times Q}}{D_{1\times 1}} \tag{27}$$

6 | RESULTS AND DISCUSSIONS

6.1 | Experimental setup

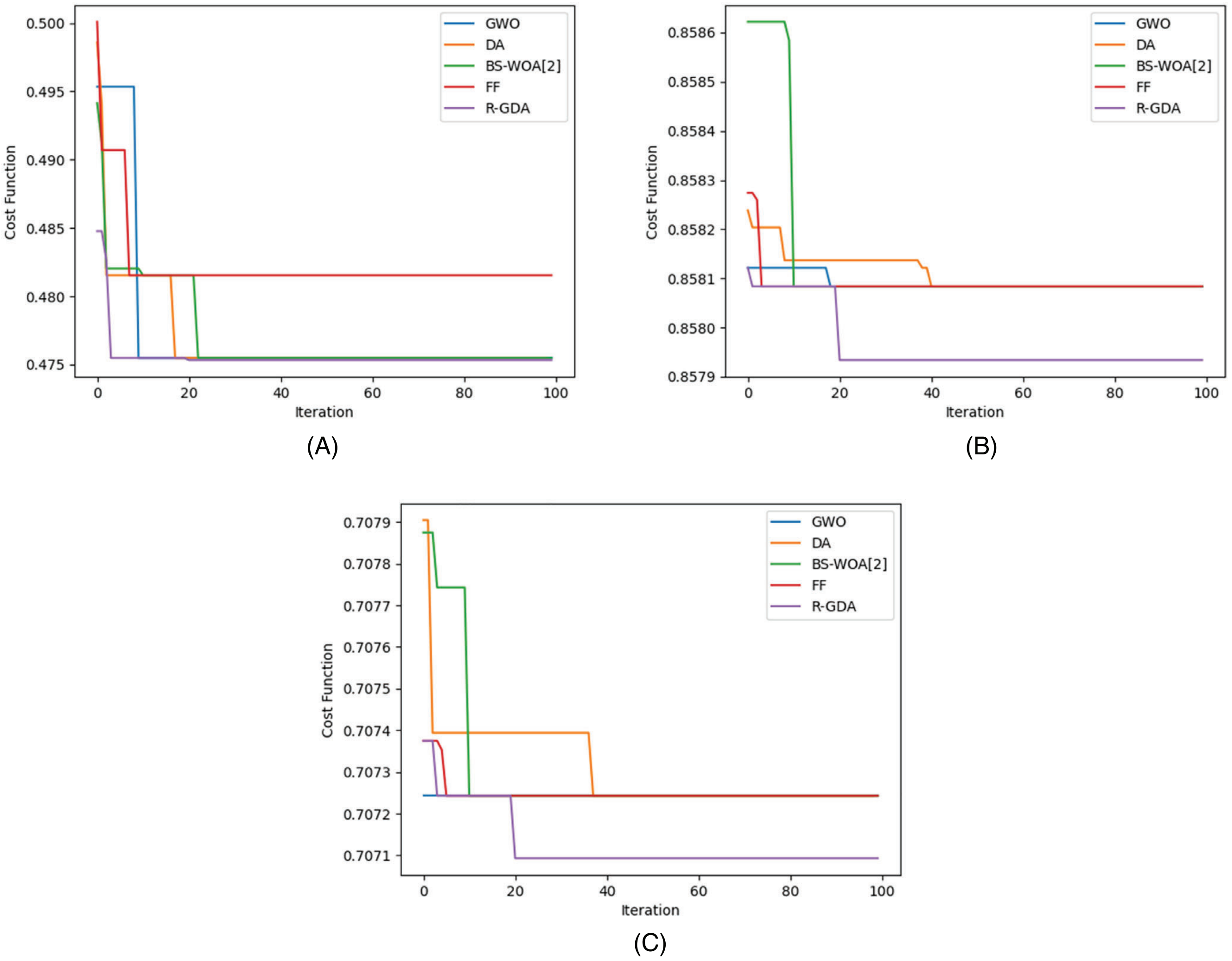
The adopted model for privacy preservation using R-GDA was executed in MATLAB and the outcomes were achieved. Consequently, the superiority of the presented model was validated by comparing it with other existing schemes like GWO,[27] DA,[30] BS-WOA,[19] and FF.[34] In addition, the analysis was carried out with respect to privacy, utility, sanitization effectiveness and restoration effectiveness. In addition, convergence analysis was performed along with attack analysis. Accordingly, the analysis was held by varying the learning percentages that ranges from 50, 60, 70, 80, 90, and 100.

6.2 | Dataset description

Chess, T1014D100K and retail datasets are the three datasets used here. The used dataset is downloaded from Reference 35. Roberto Bayardo created the chess database used in the analysis while the IBM research department created the T10I4D100 K database. The retail dataset was given by the author Tom Brijs based on data acquired from the Belgian retail store.

6.3 | Convergence analysis

The convergence analysis (cost function) for adopted R-GDA model over conventional schemes like GWO, DA, BS-WOA, and FF for datasets 1, 2, and 3 are given by Figure 3 for varied iterations that ranges from 0, 20, 40, 60, 80, and 100. From the analysis, the presented model has revealed a minimal cost value for all datasets over other compared models, thus ensuring the enhanced performance of the developed model. Particularly, from Figure 3A, a minimal cost of $0.476 has been attained by R-GDA scheme for dataset 1, which is 0.21%, 0.21%, 0.21%, and 1.47% superior to traditional GWO, DA, BS-WOA, and FF models at 80th iteration. Likewise, the cost analysis for the proposed R-GDA model over conventional schemes for dataset 2 and dataset 3 has exhibited minimal cost values for all iterations. Thus, from the evaluation, it is obvious that the presented R-GDA model has attained better performance in cost evaluation when compared to the state-of-art techniques.



(A)

(B)

(C)

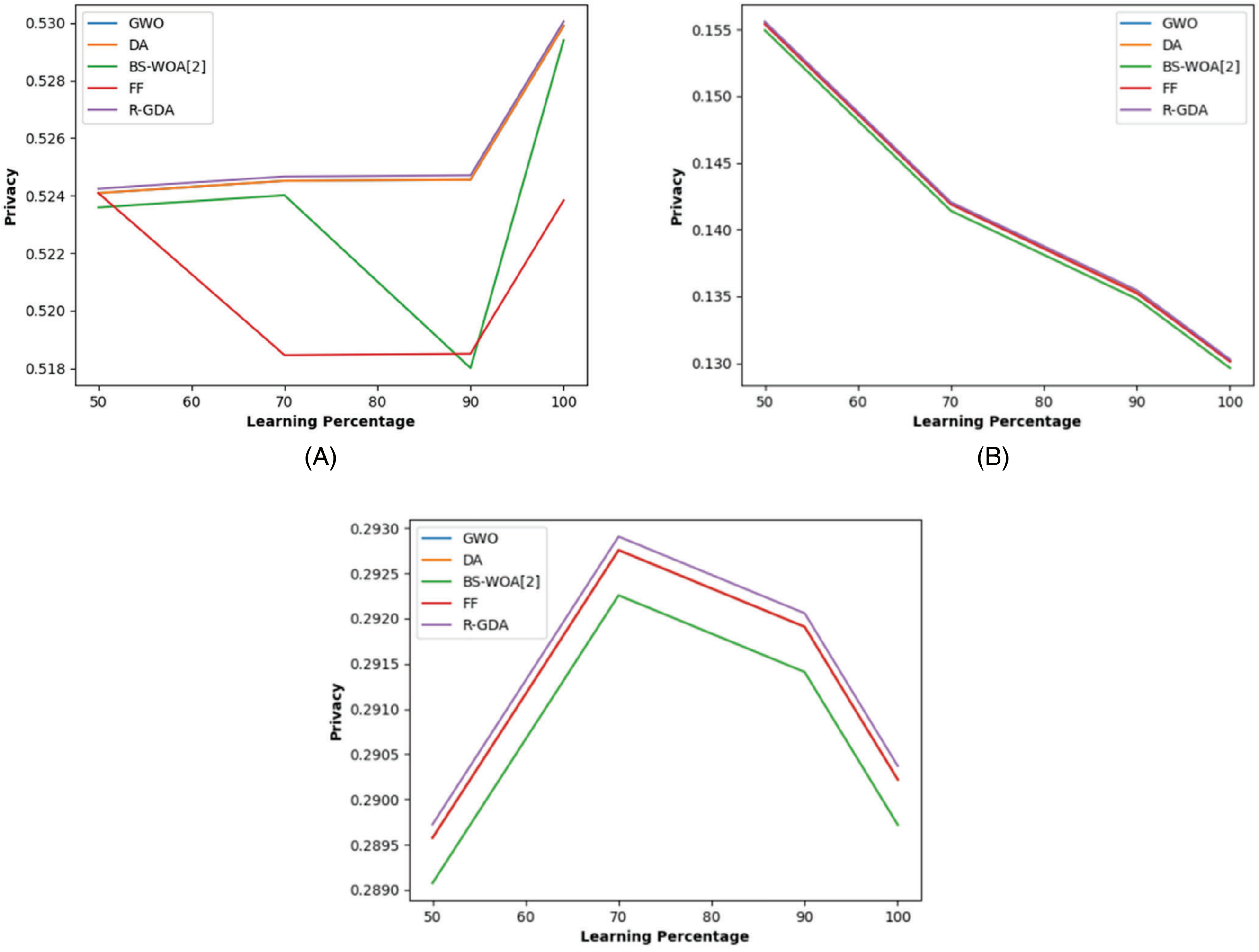
FIGURE 3 Convergence analysis attained by developed model over existing models for (A) dataset 1, (B) dataset 2, and (C) dataset 3

6.4 | Performance analysis on privacy

The performance of R-GDA model over existing models such as GWO, DA, BS-WOA, and FF is specified in Figure 4 for varied learning rates that range from 50, 60, 70, 80, 90, and 100. Accordingly, Figure 4A,B,C demonstrates the privacy outcomes attained using dataset 1, dataset 2, and dataset 3, respectively. From the analysis, the presented model has attained better values for privacy when compared over the other existing models. From Figure 4A, the R-GDA is 0.21%, 0.21%, 0.25%, and 1.13% better than the traditional models like GWO, DA, BS-WOA, and FF, respectively, at 100th learning percentage. Thus, from the evaluation, it is obvious that the privacy of the proposed R-GDA model is enhanced over the traditional ones. On examining Figure 4B,C, the privacy values goes on decreasing with increase in learning percentages for both presented as well as existing models; however, the presented approach has accomplished higher privacy than the compared models. Therefore, from the above depiction, the R-GDA is found to be enhanced than compared ones.

6.5 | Sanitization effectiveness

Figure 5 exhibits the valuation of sanitization effectiveness for developed model over existing models with respect to varying learning percentages that ranges from 50, 70, 90, and 100. In fact, the sanitization effectiveness has to be higher for the enhanced performance of the privacy system. From Figure 5B, the R-GDA model using dataset 2 has attained the higher sanitization effectiveness of 480 at 50th learning percentage that is higher than the extant models. Particularly, from Figure 5C, at 50th learning percentage, the adopted approach for dataset 3 is 4.08%, 3.06%, 4.08%, and

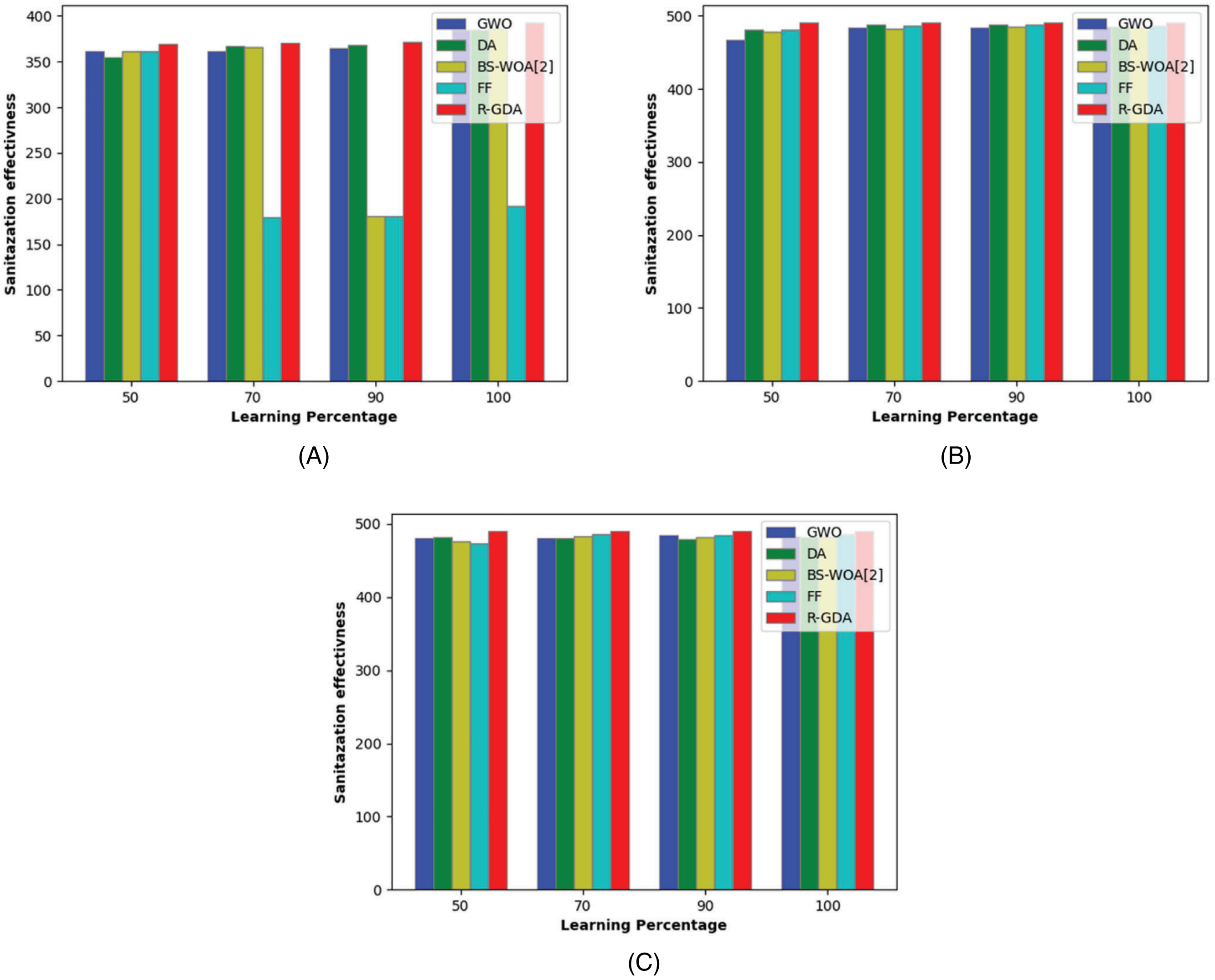


(A)

(B)

(C)

FIGURE 4 Analysis on privacy attained by developed model over existing models for (A) dataset 1, (B) dataset 2, and (C) dataset 3



(A)

(B)

(C)

FIGURE 5 Sanitization effectiveness attained by developed model over existing models for (A) dataset 1, (B) dataset 2, and (C) dataset 3

5.1% better than the traditional models like GWO, DA, BS-WOA, and FF, respectively. Thus, from the evaluation, it is clear that the R-GDA model exhibits higher sanitization effectiveness compared to the conventional models.

6.6 | Analysis on utility

The utility values attained by developed model over the existing ones are portrayed in Figure 6 for dataset 1, dataset 2, and dataset 3. Here, the utility values of R-GDA are higher when compared to the extant models for all learning percentages. In Figure 6A, the R-GDA for dataset 1 is 30.56%, 2.17%, 74.47%, and 74.47% better than the traditional models like GWO, DA, BS-WOA, and FF, respectively, at 90th learning percentage. Thus, from the evaluation, it is clear that the presented R-GDA model has obtained improved utility than existing techniques.

6.7 | Restoration effectiveness

Table 2 describes the analysis on restoration effectiveness attained by adopted R-GDA scheme over traditional schemes for varying learning percentages that ranges from 50, 70, 90, and 100. On observing the analysis outcomes, the proposed R-GDA model has attained better performance for all learning percentages when compared over the existing schemes. More particularly, on observing dataset 1 from Table 2, the adopted scheme has attained higher restoration effectiveness (0.92), and it is 67.39%, 36.96%, 67.39%, and 64.13% superior to traditional GWO, DA, BS-WOA, and

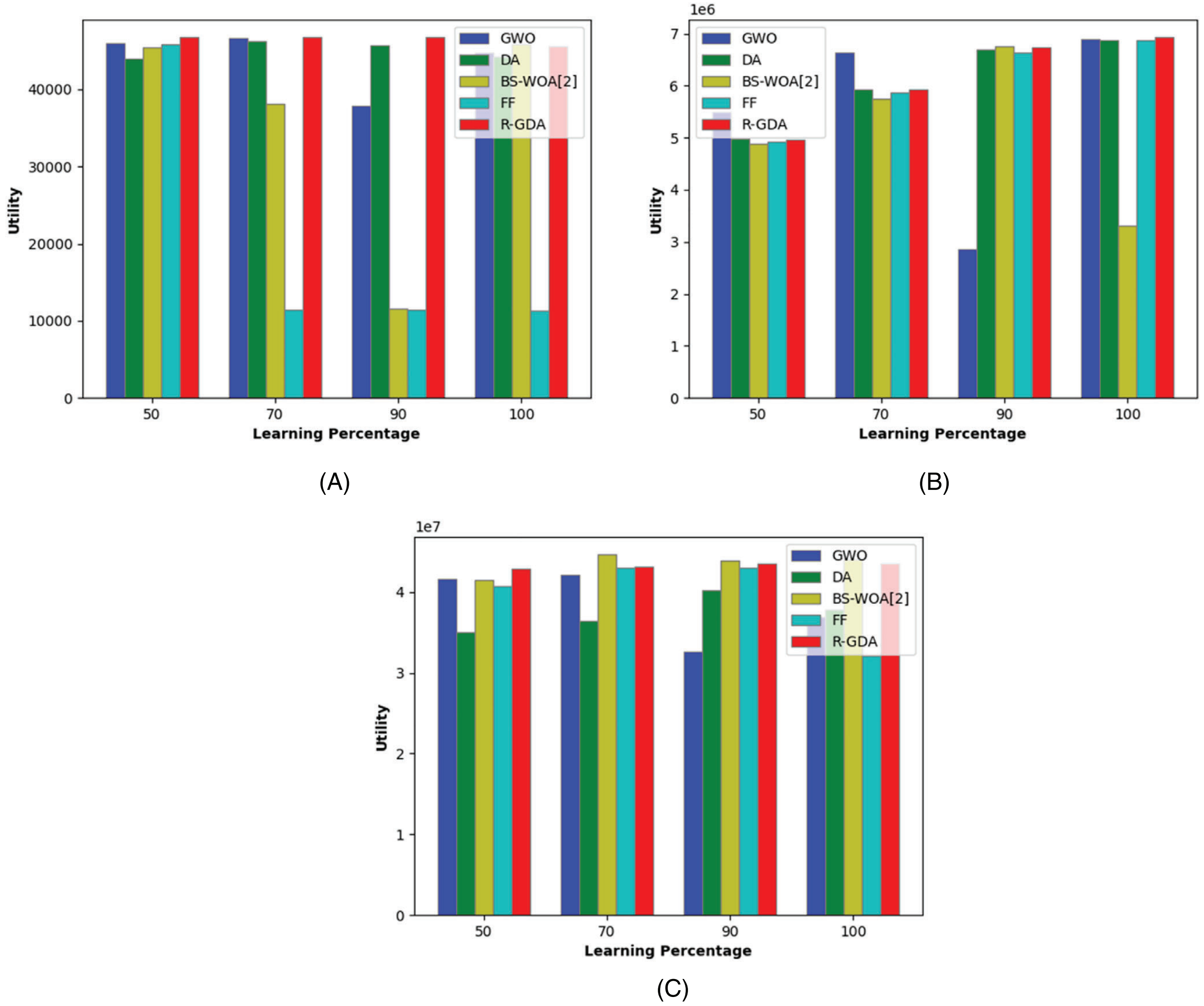


FIGURE 6 Analysis on utility attained by developed model over existing models for (A) dataset 1, (B) dataset 2, and (C) dataset 3

TABLE 2 Analysis on restoration effectiveness attained by implemented model over existing models

	Dataset 1					Dataset 2					Dataset 3				
Learning percentage	GWO	DA	BS-WOA[19]	FF	R-GDA	GWO	DA	BS-WOA[19]	FF	Prop	GWO	DA	BS-WOA[19]	FF	R-GDA
50	0.30	0.58	0.30	0.33	0.92	1	1	1	1	1	1.00	1.00	1.00	1.00	1.00
70	0.55	0.30	0.30	0.30	0.92	1	1	1	0.999999	1	1.00	1.00	1.00	1.00	1.00
90	0.33	0.30	0.30	0.30	0.92	1	1	1	1	1	1.00	1.00	1.00	1.00	1.00
100	0.55	0.63	0.63	0.50	0.89	1	1	1	1	1	1.00	1.00	1.00	1.00	1.00

FF models at 50th learning percentage. In addition, the developed model using dataset 2 has attained the best restoration effectiveness than the existing models for all learning percentages. Thereby, the overall assessment shows the betterment of developed model.

6.8 | Analysis on attacks

The analysis on attacks such as KPA and KCA attained by developed model and conventional models for varied learning percentages is elaborated by Tables 3 and 4. Table 3 reveals the analysis on KPA attack, whereas Table 4 reveals the analysis on CPA attack.

TABLE 3 Computation on KPA attack attained by implemented model over existing models

	Dataset 1					Dataset 2					Dataset 3				
Learning percentage	GWO	DA	WOA	FF	R-GDA	GWO	DA	BS-WOA[19]	FF	R-GDA	GWO	DA	BS-WOA[19]	FF	R-GDA
50	0.98	0.98	0.98	0.98	0.98	−0.26901	−0.26901	−0.26901	−0.26684	−0.26901	0.9426	0.9426	0.9426	0.9426	0.9426
70	0.98	0.98	0.98	0.98	0.98	−0.33743	−0.33743	−0.33743	−0.33743	−0.33743	−0.78572	−0.78572	−0.78572	−0.78572	−0.78572
90	0.98	0.98	0.98	0.98	0.98	−0.74398	−0.74398	−0.74398	−0.74385	−0.74398	0.89418	0.89418	0.89418	0.89418	0.89418
100	0.98	0.98	0.98	0.98	0.98	0.847102	0.847102	0.847102	0.847102	0.847102	0.844441	0.844441	0.844441	0.844441	0.844441

TABLE 4 Computation on CPA attack attained by implemented model over existing models

	Dataset 1					Dataset 2					Dataset 3				
Learning	GWO	DA	WOA	FF	R-GDA	GWO	DA	BS-WOA[19]	FF	R-GDA	GWO	DA	BS-WOA[19]	FF	R-GDA
50	0.976174	0.976174	0.976174	0.976174	0.976174	−0.24993	−0.24993	−0.24993	−0.27323	−0.24993	0.943809	0.943809	0.943809	0.943809	0.943809
70	0.976174	0.976174	0.976174	0.976174	0.976174	−0.33451	−0.33451	−0.33451	−0.33451	−0.33451	−0.78578	−0.78578	−0.66691	−0.78578	−0.78578
90	0.976174	0.976174	0.976174	0.976174	0.976174	−0.76873	−0.76873	−0.76873	−0.73358	−0.76873	−0.04243	0.894982	0.894982	0.894982	0.894982
100	0.976956	0.876596	0.976956	0.976956	0.976956	0.825459	0.825459	0.825459	0.825459	0.825459	0.849282	0.849282	0.849282	0.849282	0.849282

7 DISCUSSION

To evaluate the parameters on the R-GDA, experiments have been carried out with three databases. The entire three databases are evaluated for the Privacy, Utility, Sanitization effectiveness, and Restoration effectiveness. The privacy outcome attains through dataset 1, dataset 2, and dataset 3, whereas the developed model is compared with the other existing models like GWO, DA, BS-WOA, and FF. The privacy of the proposed R-GDA model is enhanced over the traditional ones. The sanitization effectiveness must be higher for the enhanced performance of the privacy system. The R-GDA model using dataset 2 has attained the higher sanitization effectiveness than the existing models. For all the three datasets, the presented R-GDA model has obtained improved utility than other traditional schemes. For the developed method, the restoration effectiveness must be high and the restoration effectiveness is better for the dataset 2. Thus, the overall performance provides the superior performance of developed model.

8 CONCLUSION

This article has established an optimization-based privacy preservation model via selecting the optimal key matrix. Accordingly, privacy preservation was done under data sanitization and restoration phases. During data sanitization, the data (message) were preserved by means of the optimal key. Here, the key matrix is optimally chosen using a novel hybrid algorithm. Finally, data restoration takes place that was the inverse process of sanitization. For optimization purpose, this work deployed a hybrid optimization approach known as R-GDA. In the end, the superiority of developed approach was validated over existing approaches in terms of different measures. On observing the analysis outcomes, the presented R-GDA model in terms of privacy was 0.21%, 0.21%, 0.25%, and 1.13% better than the traditional models like GWO, DA, BS-WOA, and FF, respectively, at 100th learning percentage. In addition, the R-GDA for dataset 1 was 30.56%, 2.17%, 74.47%, and 74.47% better than the traditional models like GWO, DA, BS-WOA, and FF, respectively, at 90th learning percentage. Thus, the effectiveness of the presented model was proved from the simulated outcomes. For future work, the study will focus on preserving other forms of sensitive knowledge, such as frequent itemset.

NOMENCLATURE

ANN	artificial neural network
ACL	access control list
AKPS	attribute-keyword-based data publish-subscribe
BSO	brain storm optimization
BS-WOA	brain storm-based whale optimization algorithm
DA	dragonfly algorithm
ECC	elliptic curve cryptography
EHR	e-health records
FCM	fuzzy C means
FF	FireFly
GGWO	genetic grey wolf optimization
GWO	grey wolf optimization
R-GDA	random-based grey dragon algorithm
TSP	Tracy-Singh product
WOA	whale optimization algorithm

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Frequent Itemset Mining Dataset at http://fimi.uantwerpen.be/data/.[35]

ORCID

Kingsleen Solomon Doss https://orcid.org/0000-0002-0732-3481

REFERENCES

1. Mandala J, Rao MC. Privacy preservation of data using crow search with adaptive awareness probability. *J Inf Secur Appl*. 2019;44:157-169.
2. Upadhyay S, Sharma C, Sharma P, Bharadwaj P, Seeja KR. Privacy preserving data mining with 3-D rotation transformation. *J King Saud Univ Comput Inf Sci*. 2018;30:524-530.
3. Sui P, Li X. A privacy-preserving approach for multimodal transaction data integrated analysis. *Neurocomputing*. 2017;253:56-64.
4. Wei R, Tian H, Shen H. Improving k-anonymity based privacy preservation for collaborative filtering. *Comput Electr Eng*. 2018;67:509-519.
5. Sánchez D, Batet M. Privacy-preserving data outsourcing in the cloud via semantic data splitting. *Comput Commun*. 2017;110:187-201.

6. Li CT, Shih DH, Wang CC. Cloud-assisted mutual authentication and privacy preservation protocol for telecare medical information systems. *Comput Methods Prog Biomed*. 2018;157:191-203.
7. Zhang K, Liang X, Baura M, Lu R, Shen X. PHDA: a priority based health data aggregation with privacy preservation for cloud assisted WBANs. *Inf Sci*. 2014;284:130-141.
8. Wang W, Chen L, Zhang Q. Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation. *Comput Netw*. 2015;88:136-148.
9. Hashi Y, Uchibayashi T, Hidano S, et al. Data protection for cross-border live migration in multi-cloud environment. Proceedings of the 2016 Fourth International Symposium on Computing and Networking (CANDAR), Hiroshima; 2016:681-685.
10. Kaaniche N, Laurent M. Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms. *Comput Commun*. 2017;111:120-141.
11. Shen H, Zhang M, Wang H, Guo F, Susilo W. A cloud-aided privacy-preserving multi-dimensional data comparison protocol. *Inf Sci*. 2021;545:739-752.
12. Li J, Wei J, Liu W, Hu X. PMDP: a framework for preserving multiparty data privacy in cloud computing. *Secur Commun Netw*. 2017;2017:1-14.
13. Guan Z, Zhang Y, Longfei W, Jun W, Hu J. APPA: an anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT. *J Netw Comput Appl*. 2019;125:82-92.
14. Xu X, Fu S, Qi L, et al. An IoT-oriented data placement method with privacy preservation in cloud environment. *J Netw Comput Appl*. 2018;124:148-157.
15. Fang C, Guo Y, Wang N, Ju A. Highly efficient federated learning with strong privacy preservation in cloud computing. *Comput Secur*. 2020;96:101889.
16. Prabha KM, Saraswathi V. Suppressed K-anonymity multi-factor authentication based Schmidt-Samoa cryptography for privacy preserved data access in cloud computing. *Comput Commun*. 2020;158:85-94.
17. Marsaline Beno M, Valarmathi IR, Swamy SM, Rajakumar BR. Threshold prediction for segmenting tumour from brain MRI scans. *Int J Imaging Syst Technol*. 2014;24(2):129-137. doi:10.1002/ima.22087
18. Piao C, Shi Y, Yan J, Zhang C, Liu L. Privacy-preserving governmental data publishing: a fog-computing-based differential privacy approach. *Future Gener Comput Syst*. 2019;90:158-174.
19. Thanga Revathi S, Ramaraj N, Chithra S. Brain storm-based whale optimization algorithm for privacy-protected data publishing in cloud computing. *Clust Comput*. 2019;22:3521-3530. doi:10.1007/s10586-018-2200-5
20. Yang K, Zhang K, Jia X, Hasan MA, Shen XS. Privacy-preserving attribute-keyword based data publish-subscribe service on cloud platforms. *Inf Sci*. 2017;387:116-131.
21. Majeed A. Attribute-centric anonymization scheme for improving user privacy and utility of publishing e-health data. *J King Saud Univ Comput Inf Sci*. 2019;31(4):426-435.
22. Shani Raj B, Kumar A, Venkatesan GKD. A security-attribute-based access control along with user revocation for shared data in multi-owner cloud system. *Inf Secur J: A Global Perspective*. 2020;30:309-324. doi:10.1080/19393555.2020.1842568
23. Belguith S, Kaaniche N, Laurent M, Jemai A, Attia R. Accountable privacy preserving attribute based framework for authenticated encrypted access in clouds. *J Parallel Distrib Comput*. 2020;135:1-20.
24. Sudhakar RV, Rao TCM. Security aware index based quasi-identifier approach for privacy preservation of data sets for cloud applications. *Clust Comput*. 2020;23:2579-2589. doi:10.1007/s10586-019-03028-7
25. Bibal Benifa JV, Venifa Mini G. Privacy based data publishing model for cloud computing environment. *Wirel Pers Commun*. 2020;113:2215-2241. doi:10.1007/s11277-020-07320-3
26. John AA, Deepajothi S. Privacy preservation of data sets in data mining. *Int J Eng Res Technol*. 2013;2:2278-0181.
27. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Adv Eng Softw*. 2014;69:46-61.
28. Langville AN, Stewart WJ. The Kronecker product and stochastic automata networks. *J Comput Appl Math*. 2004;167:429-447.
29. Liu S, Li T. A new hypernetwork model based on matrix operation. Proceedings of the 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE); Taipei; 2015:176-182.
30. Mirjalili S. Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Comput & Applic*. 2016;27:1053-1073.
31. Thomas R, Rangachar MJ. Hybrid optimization based DBN for face recognition using low-resolution images. *Multimed Res*. 2018;1(1):33-43.
32. Devagnanam J, Elango NM. Optimal resource allocation of cluster using hybrid grey wolf and cuckoo search algorithm in cloud computing. *J Netw Commun Syst*. 2020;3(1):31-40.
33. Shareef SKM, Rao RS. A hybrid learning algorithm for optimal reactive power dispatch under unbalanced conditions. *J Comput Mech Power Syst Control*. 2018;1(1):26-33.
34. Fister I, Fister I, Yang X-S, Brest J. A comprehensive review of firefly algorithms. *Swarm Evol Comput*. 2013;13:34-46.
35. http://fimi.uantwerpen.be/data/

How to cite this article: Solomon Doss K, Kamalakkannan S. Hybrid optimization-based privacy preservation of database publishing in cloud environment. *Concurrency Computat Pract Exper*. 2022;e6844. doi: 10.1002/cpe.6844