symmetry

MDPI

Article

Enhancement of an Optimized Key for Database Sanitization to Ensure the Security and Privacy of an Autism Dataset

Md. Mokhlesur Rahman [1,*], Ravie Chandren Muniyandi [1,*], Shahnorbanun Sahran [2] and Suziyani Mohamed [3]

[1] Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600 UKM, Selangor, Malaysia
[2] Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600 UKM, Selangor, Malaysia; shahnorbanun@ukm.edu.my
[3] Centre of Community Education & Well-Being, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600 UKM, Selangor, Malaysia; suziyani@ukm.edu.my
* Correspondence: mmarks_cse@yahoo.com (M.M.R.); ravie@ukm.edu.my (R.C.M.)

Abstract: Interrupting, altering, or stealing autism-related sensitive data by cyber attackers is a lucrative business which is increasing in prevalence on a daily basis. Enhancing the security and privacy of autism data while adhering to the symmetric encryption concept is a critical challenge in the field of information security. To identify autism perfectly and for its data protection, the security and privacy of these data are pivotal concerns when transmitting information over the Internet. Consequently, researchers utilize software or hardware disk encryption, data backup, Data Encryption Standard (DES), TripleDES, Advanced Encryption Standard (AES), Rivest Cipher 4 (RC4), and others. Moreover, several studies employ k-anonymity and query to address security concerns, but these necessitate a significant amount of time and computational resources. Here, we proposed the sanitization approach for autism data security and privacy. During this sanitization process, sensitive data are concealed, which avoids the leakage of sensitive information. An optimal key was generated based on our improved meta-heuristic algorithmic framework called Enhanced Combined PSO-GWO (Particle Swarm Optimization-Grey Wolf Optimization) framework. Finally, we compared our simulation results with traditional algorithms, and it achieved increased output effectively. Therefore, this finding shows that data security and privacy in autism can be improved by enhancing an optimal key used in the data sanitization process to prevent unauthorized access to and misuse of data.

Keywords: autism dataset; autism spectrum disorder; data sanitization; enhanced combined PSO-GWO; optimized key generation; security and privacy

Citation: Rahman, M.M.; Muniyandi, R.C.; Sahran, S.; Mohamed, S. Enhancement of an Optimized Key for Database Sanitization to Ensure the Security and Privacy of an Autism Dataset. *Symmetry* **2021**, *13*, 1912. https://doi.org/10.3390/sym13101912

Academic Editor: Alexander Shelupanov

Received: 31 August 2021
Accepted: 5 October 2021
Published: 11 October 2021

1. Introduction

Diagnostic and Statistical Manual of Mental Disorders 5th ed. (DSM-5) [1,2] defined autism spectrum disorder (ASD) as persistent deficits in two areas of development, namely social communication as well as restricted and repetitive behaviors. Children with ASD have a distinct set of deficits but with different levels of severity. Because of this, DSM-5 has divided ASD into three levels of severity based on the support required by children with ASD in their daily lives. These severity levels range from level one to level three, and are known as requiring support, requiring substantial support, and requiring very substantial support, respectively. Children with ASD demonstrate poor social communication skills, as they have deficits in verbal communication, non-verbal communication, and social-emotional reciprocity. Deficits in verbal communication cause children with ASD to exhibit difficulties in understanding spoken language and the use of inappropriate tone of voice during conversation.

In comparison, deficits in non-verbal communication lead to difficulties in understanding the meaning of body gestures, avoidance of eye contact, and inappropriate facial expressions. Furthermore, children with ASD have difficulty deciding when and how to use these nonverbal communication cues. Deficits in social-emotional reciprocity are manifested by difficulties in recognizing their own emotions, expressing their own emotions, recognizing others' emotions, easily feeling overwhelmed when they are in social situations, and difficulty taking turns during a conversation. Children with ASD also demonstrate restricted and repetitive patterns of behavior, interest, or activities. Deficits in this area are characterized by repetitive body movements and motions, ritualistic behaviors, restricted or extreme interest in specific activities or objects, and obsession with their routine. In this work, information security, a critical aspect of the symmetry concept, has been emphasized because it is very important to protect data for children with ASD. Information security refers to a set of procedures and techniques designed and implemented to prevent unauthorized access, disruption, abuse, or the alteration of private, confidential, and sensitive information or data [3]. It is an urgent desire in healthcare data [4]. A sanitizing technique has been implemented in the domain of autism data for information security.

However, the crucial issue in the transmission of autism data is security, privacy, and accuracy, which are more essential for a model or a framework, because many medical data are gathered daily [5,6]. Security and privacy are not the same issues at all. Defensive digital security precautions are employed to halt unlawful access to databases, which are known as data security. It includes integrity and availability and emphasizes defending data from malicious attacks or stealing data to make a profit. Moreover, data privacy implies dealing with the capability of a person or an organization to control which types of data through the transmission line should be exchanged, among others. Therefore, it is considered an important issue related to information sharing. Security is vital for protecting data, but it is inadequate to address without privacy.

There are different categories of data related to autism: clinical and screening [7]. Broadly, these data are medical data (clinical data) involved in health-related information consisting of regular patient care or a clinical trial program partly. An electronic health record is the best form of clinical data, the digital version of a patient's medical information and history. In addition, electronic versions of record-keeping can ensure the efficiency of coordination and sharing information between sectors such as health, education, and social care [8,9]. For example, physiological data on ASD conditions include historical information, specific characteristics, degrees of severity, and associated medical and mental health conditions [10]. They can also include information about the age at diagnosis, timely monitoring results, and the status of medication prescriptions. Other forms of data are on related intervention and support received by children with ASD. These data consist of speech therapy, occupational therapy, applied behavior analysis (ABA), early intervention, and educational programs.

Moreover, these data can also include details on the frequency and cost of therapy, an Individual Education Plan (IEP), and an Individualized Family Service Program (IFSP). Healthcare data are sensitive data that require additional protection because they come from within a person's most intimate sphere. Unauthorized disclosure may lead to discrimination and violations of fundamental rights [11,12]. For example, these data could be exploited, misused, or misinterpreted for a certain purpose [10,13]. Security or authentication systems are often required for an individual that processes and stores medical records [14,15]. In addition, many countries have developed standards for a doctor–patient relationship that preserve confidentiality [16]. These standards protect patients' dignity and ensure that patients provide accurate information to receive the correct treatment.

There is a growing trend of hacking into healthcare data, because they are an attractive target. Healthcare systems are an easy target for hackers due to their interconnectedness, easily accessible access points, outdated systems, and a lack of emphasis on cybersecurity [13]. Hackers steal health information because they can make money from it [15]. A single patient file can be sold for a hundred USD [17], and a complete set of medical creden-

tials can be sold for over a thousand USD [13,18]. Personal information stated in healthcare records can be used for opening bank accounts, securing loans, or getting a passport [19]. Deficits in social communication and behavior, as stated in medical information, mean that the buyers of information can easily act as a person with ASD. They can also use this disability issue to escape from inconvenient situations, and it is hard for authorities to detect them. Due to this circumstance, healthcare data security is a crucial issue.

Various frameworks or models use different techniques, methods, or algorithms that attain accuracy, data security, and privacy issues, such as cost-effective and model-driven application-level frameworks for e-health data transmission using different encrypted and decrypted algorithms, namely, DES, 3DES or TripleDES, AES, Blowfish, IDEA, and RC4. Some researchers utilize various meta-heuristics algorithms such as artificial bee colony (ABC) [5], particle swarm optimization (PSO) [20], crow search algorithm (CSA) [21], glowworm swarm optimization (GSO) [22], grey wolf optimizer (GWO) [23], and others. To address the security and privacy problems, some of these investigations use k-anonymity and query. Such approaches need a large amount of time and computer resources. In addition, some of these traditional meta-heuristics algorithms also possess lower solving precision, slower convergence, and worse local searching ability. Moreover, we identified certain critical issues in the existing studies [5,6,24–26] which we addressed, thus forming the focus of our research contributions. Such critical issues, put in question form, include, but are not limited to, the following:

- For how long will the key value be updated during the key generation stage?
- The key length will be allocated based on which value?
- How are the values of the parameters defined?
- What is the key range value?

In addressing the above issues, we applied data sanitization for autism data security for better accuracy, security, and privacy. Data sanitization is a process that disguises sensitive information in order to facilitate database testing and development [27]. This can be done by overwriting it with similar types of false data while looking realistic. It is essential to protect vulnerable information, and there is an ethical obligation to do that in many countries. There are various data sanitization techniques, such as encryption/decryption, gibberish generation, number variance, shuffling records, substitution, masking data, and NULL'ing Out. We applied an optimal key that is utilized in the data sanitization technique.

However, the objectives of this study can be summarized as below:

- First, to propose a data sanitization process.
- Secondly, to enhance an optimal key by considering the above issues, which is used in the data sanitization procedure for the security and privacy of ASD datasets.
- Finally, to compare the accuracy achieved by our optimal key with the accuracy of other existing security and privacy frameworks.

The paper is structured as follows: Section 2 analyzes the relevant works on the application of various encryption and decryption algorithms and techniques. In Section 3, we describe the methodology. Sections 4 and 5 demonstrate the experiments, results, and discussions, respectively, for ASD datasets, including possible solutions. Finally, we conclude this work in Section 6, along with the future direction.

2. Related Works

This section reviewed and analyzed relevant works on sensitive data security and privacy, as well as concerns that must be addressed, and summarized the aspects and challenges of various security and privacy models in Table 1.

2.1. Security and Privacy in Processing Medical Data

Mewada S et al. [5] used an artificial bee colony-based (ABC-based) model to create a privacy model for hiding sensitive information in medical data. The ABC-based model creates an optimal key for anonymizing sensitive information, and the same key was used

for restoring information. They also considered four threats, for example, known cipher attack (KCA), known-plaintext attack (KPA), chosen cipher attack (CCA), and chosen-plaintext attack (CPA), for the validity of the performance of their suggested approach.

Depending on adaptive awareness probability with a meta-heuristic algorithm, the crow search algorithm [6] improved the data preservation method for medical data. The suggested framework deals with the method of data sanitization to mask sensitive laws. In comparison to other existing techniques, the effectiveness of the suggested system was observed, and it was found that their suggested system offers rigorous and efficient results for the security of autism data.

To select and classify autism spectrum disorder (ASD), Rahman MM et al. [7] reviewed state-of-the-art articles. After reviewing the works, they emphasized data security and privacy in order to identify autistic features perfectly and quickly.

Data security and privacy are also significant concerns for the cloud computing environment, because this environment provides access to various data, files, and applications. Due to its advantages, the cloud is widely exploited in the healthcare sector. For example, in work [22], Alphonsa MMA et al. developed a secure model named GMGW to sanitize sensitive information of heart disease data based on the cloud system. In the same way, the authors in [24,25,28] developed security and privacy models individually by applying different algorithms to the cloud computing system for data security and compared the performance of their models with the conventional algorithms for more improvement.

Abidi MH et al. [26] established a secured data transmission model as Whale with New Crosspoint-based Update (WNU) in supply chain management along with blockchain technology. They also evaluated their model concerning four research issues, namely false rule generation (FR), the information preservation (IP) rate, the hiding failure (HF) rate, and the degree of modification (DM). Ochôa IS et al. [29] also applied blockchain technology to protect users' personal data by using three blockchains to confirm security, trust, and privacy in their architecture. They utilized sidechains for scalability and adaptability of their system.

Shailaja GK et al. [30] applied an optimal key in their proposed model for the privacy-preserving data mining (PPDM) technique using an opposition intensity-based cuckoo search algorithm. They also assessed their model with FR, IP, HF, and DM.

In the works of [31–34], authors built security and privacy models independently by applying various algorithms on the cloud computing system. They also compared the performance of their models with the conventional approaches for enhancement.

Liu Y et al. [35] introduced a new reversible data hiding strategy based on the region of interest (ROI) in encrypted medical images. A data owner primarily divides an original diagnostic image into the region of interest (ROI) and the region of non-interest (RONI). The encryption key was subsequently used in anonymizing the images in this regard. The least significant bits (LSB) of the encrypted ROI and electronic patient record (EPR) were concatenated by a data hider. Afterward, the concatenated data were embedded into the encrypted image by the LSB substitution technique. With the data-hiding key, the receiver retrieves the embedded data contained in the encrypted medical image. If the recipient possesses the encryption key, directly decrypting the encrypted medical image could result in a medical image that is similar to the original image. But suppose the recipient had both keys (data-hiding key and encryption key); embedded data could be retrieved without any mistake, subsequently meaning that the embedded data ROI could be retrieved without any flaws.

Table 1. Significant features and challenges of several security and privacy frameworks for concealing data.

Authors/Year/References	Techniques/Methods	Attributes/Characteristics/Features	Challenges
Mewada et al. (2020) [5]	ABC algorithm	■ Anonymize sensitive information. ■ Sanitization, restoration process.	■ Time and space dimensions of privacy-preserving.
Mandala et al. (2018) [6]	AAP-CSA algorithm	■ Healthcare data sanitization. ■ Sanitization, restoration process.	■ Working with advanced meta-heuristic algorithms. ■ Considering more relevant constraints in objective functions. ■ The accuracy of the sanitization process seems not high.
Alphonsa et al. (2018) [22]	GMGW algorithm	■ Preserve sensitive healthcare data. ■ Hybridization of GA along with the GSO algorithm. ■ Analyze the effectiveness of sanitization, restoration, analysis on convergence, and key sensitivity statistically. ■ Running parallel computation is simpler. ■ Possess higher probability and proficiency in achieving the global optima.	■ Vulnerable unencrypted data gathered at the remote cloud storage server. ■ Internal and external threats launched by unreliable cloud service providers and suppliers. ■ It can converge prematurely and be trapped into a local minimum, especially with complex problems.
Ahamad et al. (2020) [24]	J-SSO algorithm	■ Usage of the beneficial features of JA and SSO algorithm collectively. ■ Work with various datasets, such as air quality, concrete data, heart disease, superconductivity, and wholesale customer data. ■ Efficient cloud data privacy preservation model with the data sanitization and restoration approach. ■ Work with multi-objective functions involving various parameters such as the degree of modification, hiding ratio, and information preservation ratio.	■ The inaccurate and inefficient offering of security measures for data transmissions and operations in the cloud. ■ Susceptible data by untrustworthy cloud environment providers.
Balashunmugaraja et al. (2020) [25]	CI-LA algorithm	■ Perform multi-objective functions, including different parameters. ■ Analyze the effectiveness of sanitization, restoration, analysis on convergence, and key sensitivity statistically.	■ Hard to configure the keys accurately. ■ Network connection dependency. ■ Essential to keep updating the new software.
Abidi et al. (2021) [26]	WNU algorithm	■ Uses the features of supply chain networks depending on blockchain technology. ■ Evaluated using HF rate, IP rate, FR generation, and DM.	■ The selection of optimal key in the key extraction phase is the most significant challenge.
Lekshmy et al. (2019) [28]	ABC algorithm	■ Available users are clustered in distributed computing. ■ Among the users from each group, a user known as a helper user transmits data nominated via the service provider. Evaluated in terms of a few factors (such as clustering accuracy, processing time, and data transmission time).	■ Big data sets are not encrypted in a distributed system by using the kernel k-means algorithm for encryption.
Shailaja et al. (2019) [30]	OI-CSA algorithm	■ Gives superior runtime and scalability.	■ Essential to increase privacy-preserving data mining.
Renuga et al. (2018) [31]	GSA algorithm	■ Lower execution time, hiding failure, maximum dissimilarity value in comparison with the existing technique.	■ Possibility of malicious threats in the sensitive information gathered in the cloud.
Han et al. (2020) [32]	CloudDLP	■ Browser-based applications on cloud storage.	■ The outside enterprise in cloud services can easily unveil documents or sensitive data in images.
Revathi et al. (2018) [33]	BS-WOA algorithm	■ Involve a small number of parameters and lack of local optima entrapment for resolving clustering problems.	■ Hard to keep up with the privacy of every database.

ABC, Artificial Bee Colony; AAP-CSA, Adaptive Awareness Probability-based CSA; GMGW, Genetically Modified Glowworm Swarm Optimization; GA, Genetic Algorithm; GSO, Glowworm Swarm Optimization; J-SSO, Jaya-based Shark Smell Optimization; JA, Jaya Algorithm; SSO, Shark Smell Optimization; CI-LA, Crossover Improved-Lion Algorithm; WNU, Whale with New Crosspoint-based Update; HF, Hiding Failure rate; IP, Information Preservation rate; FR, False Rule generation; DM, Degree of Modification; OI-CSA, Opposition Intensity-based Cuckoo Search Algorithm; GSA, Gravitational Search Algorithm; CloudDLP, Cloud Data Loss Prevention; BS-WOA, Brain Storm-based Whale Optimization.

In a study by [36], Zhang Y et al. established a Privacy-Aware Smart Health (PASH) access control system. The key ingredient of their system was a large universe ciphertext-policy attribute-based encryption (CP-ABE) whose access strategy was somewhat secret. The access strategy in the encrypted s-health records (SHRs) was that the attribute values were hidden, and only the attribute names were exposed. Indeed, attribute values hold much more sensitive information than generic attribute names. Specifically, PASH conducted an effective SHR decryption test that involves a limited number of bilinear pairings. Moreover, the attributes universe could be infinitely large, and public parameters were small and constant in size. From the analysis, they claimed that PASH was completely secure in standard frameworks.

Sharma U et al. [37] recommended two parallelized methods called PGVIR and PHCR. These approaches were applied to the spark framework, which manipulates the data so that no sensitive data could be retrieved at the time of ensuring the utility of sanitized data. Taking the standard dataset through the experiment, they found that PGVIR was more scalable while PHCR ensured the dataset's quality. Sharma S et al. [38] suggested an approach that optimally reduced the side effect of the hiding process on non-sensitive data, provided a balance between knowledge and privacy, and successfully regulated the rapid increase in data volume.

Again, some recent studies by Lin Z et al. [39–41] emphasized the secured data for multiple access in the presence of the availability of the Internet of Things (IoT). They utilize state-of-the-art technologies, such as unmanned aerial vehicle (UAV), beamforming (BF) approach, satellite, and aerial-integrated network (SAIN), rate-splitting multiple access (RSMA), etc., for high-level data rate, lower latency, and data exchange reliability.

Although some research works [42–46] employed symmetry-adapted cutting-edge technologies for diagnosing different human disabilities and illnesses, maintaining accuracy and privacy without delays, there is also an imperative urge for data security and privacy.

2.2. Features and Challenges of Privacy Preservation Models

It is noted from the comprehensive literature survey that a vast number of algorithms with advanced techniques have been generated for anonymization [47–50]. It is possible to describe these algorithms as single objective, multi-objective, and restricted algorithms. These algorithms aim to retain information or data that are sensitive.

However, none of these algorithms will ensure the protection of knowledge as required for usefulness and privacy. There is, however, a need for an efficient model of anonymization to protect medical records. The latest trends have demonstrated that confidential knowledge or sensitive information is being maintained, often by meta-heuristic algorithms. The purpose of these algorithms is to produce an optimal key for the method of sanitization. These algorithms are shown to have better outcomes in comparison with conventional algorithms. Some of the studies often used k-anonymity and query to fix the privacy issues. But these techniques take a great deal of time as well as resources for computation.

Therefore, in this study, an attempt is made to establish an optimal key for protecting privacy using PSO and GWO algorithms for the sanitization process.

3. Methodology and Architecture

The goal of this study was to come up with a potential solution or remedy to an issue. Regarding this, the problem addressed was that of yielding optimal keys using the characteristics of meta-heuristics algorithms. We compared many cutting-edge solutions to the problem in order to establish the ideal solution. Accordingly, we identified a research gap regarding the formation of optimal key in those state-of-the-art solutions. We pointed out some significant issues in the introductory section, wherein existing technologies have no definite resolution to the challenges in terms of security and privacy. Consequently, we addressed these critical issues by forming the optimal key in the proper way.

To provide the solution of the problem, the following framework was utilized. Figure 1 presents the overall architecture of our proposed model, which ensures the security and privacy of autism data and maintains our expected performances.

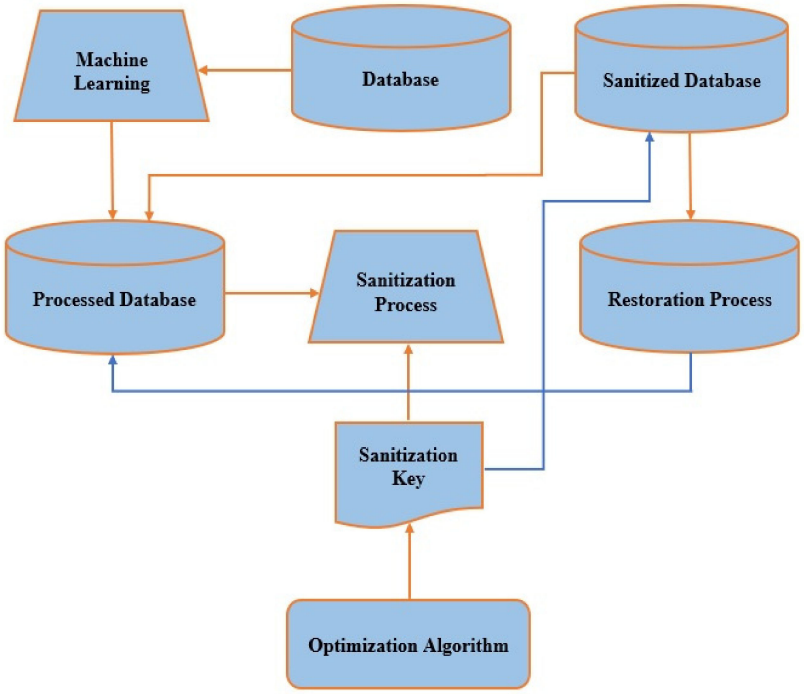

Figure 1. The main architecture of the data security and privacy model.

The different components of the architecture are as follows:

1. Original Database;
2. Machine Learning;
3. Processed Database;
4. Optimization Algorithms;
5. Sanitization Key;
6. Sanitization Process;
7. Sanitized Database;
8. Restoration Process.

In this framework, the dark orange arrows represent the sanitization process, which is the focus of this study, and the blue arrows denote the restoration process.

As a security and privacy concern, autism-related sensitive data protection was considered and implemented by means of a data sanitization technique. The major different components of the overall architecture related to the sanitizing purposes have been illustrated below, for concealing the sensitive data related to autism.

3.1. Sanitization Process

The procedure of the sanitization technique is illustrated in Figure 2. Here, D', a sanitization database, is obtained accompanied by the sanitizing key generated from the processed database during the key generation process. The resulting key matrix, K_2, and D indicate the pruned key matrix and processed database, respectively, which are binarized to fulfil the XOR function. Processed data D are obtained from the original database by using machine learning algorithms, so that no blank data, missing data, anonymous data, false data may exist. Following this binary XOR operation, the chance of having '0' is high. Getting such zero values yields insignificant data elements. So, to avoid such zero values,

a unit value (one) is added where the + (plus) sign refers to the binary summation. Then, a unit step input is summed up consequently, while D' is obtained, as shown in Equation (1).

$$D' = (K_2 \oplus D) + 1, \quad (1)$$

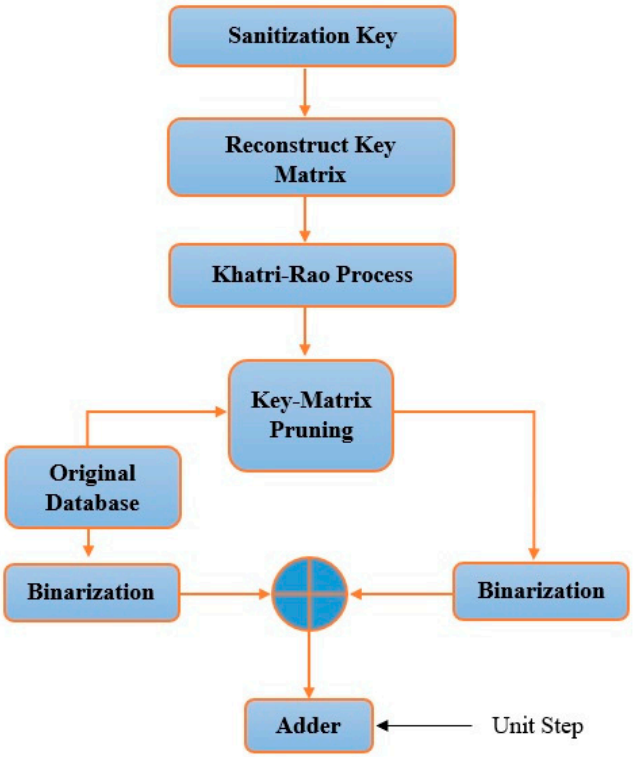

Figure 2. The architecture of the sanitization process.

3.2. Sanitization Key Generation

Figure 3 demonstrates the key generation process for sanitization purposes. The optimal key is created with the help of the proposed Enhanced Combined PSO-GWO framework by setting the population of various keys indiscriminately. It is followed by the sanitization process step, through which a sanitized database is obtained. Specifically, Figure 3 illustrates the key generation process for data sanitization and the restoration process. The proposed Enhanced Combined PSO-GWO algorithm is used at the key update step for obtaining the better key and is performed depending on an iterative loop to obtain the better solution in the process. In the interim, the sanitized database is obtained through the sanitization process. Again, the processed database acquires an association rule and measures the objective functions, C_1, C_2, and C_3, respectively. Finally, the key value is updated continuously during this process until the highest termination measure is achieved and the best-desired solution is generated. For this data sanitization process, a key is created optimally by the proposed Enhanced Combined PSO-GWO. The dimension of the chromosome is allotted depending on the value of $\sqrt{L\frac{C}{D}}$. The value fixes the elements, $\lfloor 0, \sqrt{\max(D)} \rfloor$, whereas D refers to the processed initial database.

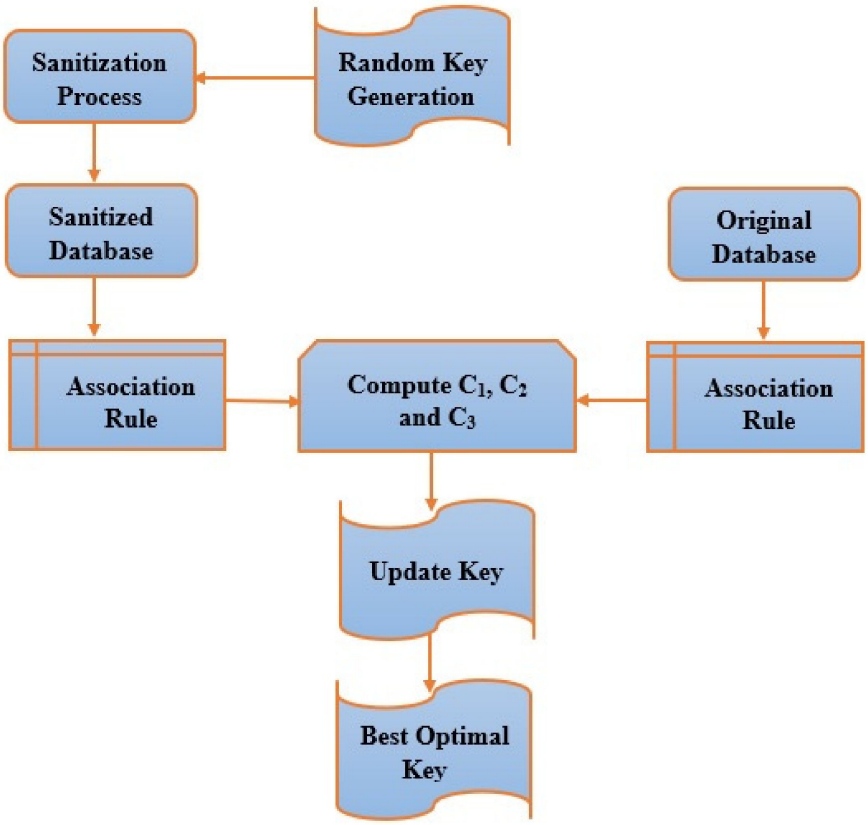

Figure 3. The architecture of the key generation process.

Procedure of Proposed Optimal Key Extraction in Sanitization Process

- Key Encoding

The usage of keys, K, for the procedure of sanitization depends on the encoding of the proposed Enhanced Combined PSO-GWO algorithm. The optimization of the number of keys ranging from key K_1 to key K_N is controlled by using an Enhanced Combined PSO-GWO algorithm, and as a result, the optimal key is obtained. The length of the key is assigned as $\sqrt{L\frac{C}{D}}$ in this case. Usually, the key length for sanitization is L_D. However, our key generation process needs $\sqrt{L\frac{C}{D}}$ and the technique of key transformation forms a key of L_D using the Khatri–Rao product. A Kronecker product that is column-wise is known as the Khatri–Rao product [51].

- Key Transformation

Let us consider a database transaction, presented in Table 2

Table 2. Data transaction in the database.

Transactions	Data		
T_1	1	2	
T_2	1	3	
T_3	2	3	4
T_4	1	3	4
T_5	3	4	

The key K is converted by applying the Khatri–Rao product during the Key Transformation process phase. This operation occurred on two matrices of arbitrary size as a block matrix and is denoted by the operator $\otimes$. From the beginning, K is mainly formed as K_1 with the dimension of the matrix, $[\sqrt{L\frac{C}{D}} \times T_{\max}]$. The recommended technique of K = 5, 0, 10, for illustration, performs row-wise duplication and produces the key matrix, K_1, with dimension $[\sqrt{L\frac{C}{D}} \times T_{\max}]$, as revealed in Equation (2), wherein the row matrix depends on $\sqrt{L\frac{C}{D}}$, as well as the column matrix, is assigned depending on T_{max}.

So, the matrix K with size, $[\sqrt{L\frac{C}{D}} \times T \max]$

$$K_1 = \begin{bmatrix} 5 & 5 & 5 \\ 0 & 0 & 0 \\ 10 & 10 & 10 \end{bmatrix}, \quad (2)$$

Similarly, by applying Khatri–Rao products like $K_1 \otimes K_1$, the key matrix, K_2, is achieved, whose dimension is $[L_D \times T \max]$. Its sizes are trimmed regarding the initial database dimensions presented in Equation (3).

$$K_2 = \begin{bmatrix} 5 & 5 & 5 \\ 0 & 0 & 0 \\ 10 & 10 & 10 \end{bmatrix} \otimes \begin{bmatrix} 5 & 5 & 5 \\ 0 & 0 & 0 \\ 10 & 10 & 10 \end{bmatrix}, \quad (3)$$

K_1 acts the key generation process depending on the Khatri–Rao approach and produces a matrix of the same size as the initial database, K_2 $[L_D \times T_{\max}]$. Finally, the rule hiding method is encompassed to obtain the sanitized database, D', by concealing the sensitive data. In addition, binarization is performed between the processed database as well as the key matrix. Consequently, the rule hiding operation is applied to the binarized key matrix pruning, wherein the XOR function takes place with the initial binarized database, accomplishing equivalent matrix sizes, and adds up with the unit value and produces the sanitized database, which is revealed in Equation (1), where K_2 implies a pruned key matrix. Furthermore, prior to the sanitization of D, D' achieved from the sanitization process, raises both sensitive rules and association rules. In this way, Equation (1) is analyzed depending on the Khatri–Rao method and is reached by sanitized database D'.

- Fitness Evaluation

The functions, C_1, C_2, and C_3, known as objective functions (hiding failure rate C_1, information preservation rate C_2, and degree of modification C_3), are assessed through Equation (4) to Equation (6) after sensitive rules and association rules of the original and sanitized database have been generated. In Equation (4), f_s and f_m refer to the frequency of the sensitive itemset, whereas f_s signifies in the case of sanitized data, and f_m implies in respect to original data. Similarly, f_{ns} represents the non-sensitive itemset frequency in reference to sanitized data shown in Equation (5). From Equation (6), the Euclidean distance is achieved where D is original data, and D' is sanitized data. Finally, the distance amidst individual items set from sanitized and original data is represented by c_4 in Equation (7). Moreover, f indicates the fitness function of the recommended method, whereas w_1, w_2, and w_3 represent the impact of a particular cost function regarding C_1, C_2, and C_3 at the same time.

$$C_1 = \frac{fs}{fm}, \quad (4)$$

$$C_2 = \frac{fns}{fm}, \quad (5)$$

$$C_3 = dist\ (D, D') \rightarrow Euclidean\ distance, \quad (6)$$

$$\mathrm{f} = w1\left(\frac{C1}{max[C1,\ C2]}\right) + w2\left(1 - \frac{C2}{max[C1,\ C2]}\right) + w3\left(\frac{C3}{max(C4)}\right), \tag{7}$$

However, the functions C_1, C_2, and C_3 are preferred to determine how efficiently the autism data are sanitized, using the recommended Enhanced Combined PSO-GWO algorithm. For medical data, the objective function of the suggested technique is presented by Equation (8).

$$G = Min\ (f) \tag{8}$$

3.3. Both Traditional PSO and GWO Algorithms

In this section, we discussed the traditional PSO algorithm in Section 3.3.1 and GWO algorithm in Section 3.3.2.

3.3.1. Traditional PSO Algorithm

In the PSO algorithm, there are three vectors. These are x-vector, p-vector, and v-vector. The x-vector keeps track of the present location for the particle in the searching area, whereas the p-vector (pbest) identifies the position of where the particle has discovered the best solution so far. Moreover, the v-vector incorporates particle velocity, indicating where every other particle will move through the following iteration. At the outset, the particles are randomly shifted in specified directions. The particle's orientation might be adjusted gradually, and as a result, it began to move in the direction of the prior best location on its own. After that, it explores the surrounding area for the best locations for some fitness functions, fit = $S^m - S$. Here, the location of the particle is provided as $\vec{M} \in S^m$, while its velocity is provided as $\vec{w}$. Initially, these two variables are picked at random and then updated repeatedly according to two formulae shown in Equation (9)

$$\vec{w} = \omega\vec{w} + c_1 r_1\left(\vec{q} - \vec{M}\right) + c_2 r_2\left(\vec{f} - \vec{M}\right), \tag{9}$$

In this case, ω, a user-defined behavioral parameter is the inertia weight, which regulates the amount of recurrence in particle velocity. The particle's previous best position (pbest position) is $\vec{q}$, and the particle's previous best position in the swarm (gbest position) is $\vec{f}$; in that way, the particles implicitly interact with each other. This is weighted using stochastic variables $r_1, r_2 \sim U\ (0, 1)$, while the acceleration constants are c_1, c_2. Regardless of fitness gains, the velocity is added to the particle's present position to propel it to the next place in the searching area, as shown in Equation (10)

$$\vec{M} \leftarrow \vec{M} + \vec{w}\ , \tag{10}$$

3.3.2. Traditional GWO Algorithm

In the GWO algorithm, there are hierarchical search agents such as level 1 (Alpha), level 2 (Beta), level 3 (Delta), and level 4 (Omega). When the grey wolves hunt their prey, then the characteristic of encircling is expressed mathematically in Equations (11) and (12).

$$\vec{B} = \left| \vec{E}.\vec{M}_q(u) - \vec{M}(u) \right|, \tag{11}$$

$$\vec{M}(u+1) = \vec{M}_q(u) - \vec{H}.\ \vec{B}, \tag{12}$$

where u is given the current iteration $\vec{H}$ and $\vec{E}$ are referred to as the coefficient vectors. Grey wolves possess a unique skill for detecting the position of prey and encircling it. These grey wolf hunting actions are mathematically reproduced utilizing alpha, beta, and delta wolves' enhanced awareness of probable prey locations. The first three best solutions are considered,

regardless of whether the remainder is required. The mathematical Equations (13)–(15) are provided below:

$$\begin{aligned} \vec{B}_{\alpha} &= \left| \vec{E}_1.\vec{M}_{\alpha} - \vec{M} \right|, \\ \vec{B}_{\beta} &= \left| \vec{E}_2.\vec{M}_{\beta} - \vec{M} \right|, \\ \vec{B}_{\delta} &= \left| \vec{E}_3 \cdot \vec{M}_{\delta} - \vec{M} \right|, \end{aligned} \tag{13}$$

$$\begin{aligned} \vec{M}_1 &= \vec{M}_{\alpha} - \vec{H}_1.\left(\vec{B}_{\alpha}\right), \\ \vec{M}_2 &= \vec{M}_{\beta} - \vec{H}_2.\left(\vec{B}_{\beta}\right), \\ \vec{M}_3 &= \vec{M}_{\delta} - \vec{H}_3.\left(\vec{B}_{\delta}\right), \end{aligned} \tag{14}$$

$$\vec{M}(u+1) = \frac{\vec{M}_1 + \vec{M}_2 + \vec{M}_3}{3}, \tag{15}$$

3.4. The Proposed Enhanced Combined PSO-GWO Algorithm

Despite having good performance, enhancements can be made to traditional algorithms to address the limitations and improve performance. The traditional PSO algorithm demonstrates a few weaknesses, such as lower performance over a wide range of fields. The GWO algorithm also has a few drawbacks: poorer local searching capability, slower convergence, and lower solving precision. Consequently, further analysis is required to improve robustness and integration.

This study implements a new hybrid algorithm for solving those issues. The proposed Enhanced Combined PSO-GWO is elaborated as follows: in this regard, the criteria of the PSO algorithm are implemented in the GWO algorithm. The enclosure of the prey mathematical model, in the suggested method, is provided in Equations (11) and (12), while the mathematical model of the hunting method is shown in Equations (13)–(15). The updating of the location is the main reformation in the suggested model. So, the updating of the location in our Enhanced Combined PSO-GWO model is shown in Equation (16), where $\vec{M}$ refers to the velocity for the updating of the location of PSO—this is demonstrated in Equations (9) and (10).

$$M(u+1) = \frac{\vec{M}_1 + \vec{M}_2 + \vec{M}_3 + \vec{M}}{4}, \tag{16}$$

Again, c_1 and c_2 are considered acceleration constants in the traditional PSO algorithm, whereas c_1 and c_2 fluctuate according to the values 0.1, 0.3, 0.5, 0.7, and 1 in the suggested Enhanced Combined PSO-GWO model. The optimal key selection based on PSO-GWO is presented in Algorithm 1.

Algorithm 1: Optimal Key Selection through Enhanced Combined PSO-GWO.

M_j is the Grey Wolf population where $j = 1, 2, N$. Here, M_α, M_β, and M_δ denote the best searching agent, 2nd best searching agent, and 3rd best searching agent, respectively. Moreover, e is the components, and H, E are coefficients. The goal of this algorithm is to output the best searching agent, M_α.
{
Set initial values to the M_j
Set initial values to e, H, and E also
Measure the fitness values of each searching agent, M_α, M_β, and M_δ.
while ($u < max$) **do**
{
for each searching agent, **do**
{
Revise the present location of the searching agents using Equation (16)
}
Revise e, H, and E
Assess fitness values for all searching agents
Revise $M\alpha$, M_β, and M_δ
u: = u + 1
}
return M_α
}

4. Experiment and Analysis

This section explains the implementation of our proposed method and types of autism datasets with sources and the compared traditional algorithms in Section 4.1. We also show our proposed methods' simulation performances compared to those conventional algorithms against various attacks in Section 4.2.

4.1. Configuration for Experiment

The proposed method was developed by using the Python programming language. The autism datasets were collected from the faculty of education, Universiti Kebangsaan Malaysia. The autism datasets applied for this study are collected from different aged-group autistic children. These include the autism child dataset at 24 months with 26 attributes and 209 instances, the autism child dataset at 30 months, which have 29 attributes and 209 instances, the autism child dataset at 36 months, including 31 attributes and 234 instances, and the autism child dataset at 48 months, including 33 attributes and 302 instances. All datasets are autism diagnostic data, which have three scoring options, such as z = 0, v = 5, and x = 10. For every type of dataset, the cut-off values were different, at 71, 95, 100, and 105, respectively. The performance of the proposed framework was compared with the existing conventional algorithms such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Crow Search Algorithm (CSA), Differential Evolution (DE), and Adaptive Awareness Probability-based CSA (AAP-CSA).

4.2. Simulation

Various types of attacks, such as the Known Cipher Attack (KCA), Known Plaintext Attack (KPA), Chosen Cipher Attack (CCA), and Chosen Plaintext Attack (CPA), were tested. Based on these attacks, the simulation was performed, as shown in Figures 4–8.

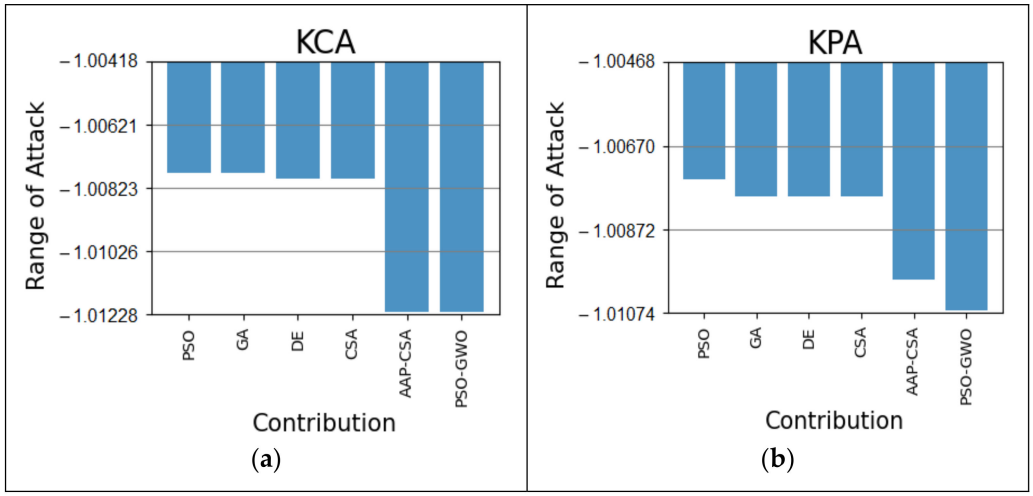

Figure 4. Analysis of the performance of various algorithms using objective functions. (**a**) Performances based on the KCA attack; (**b**) Performances based on the KPA attack.

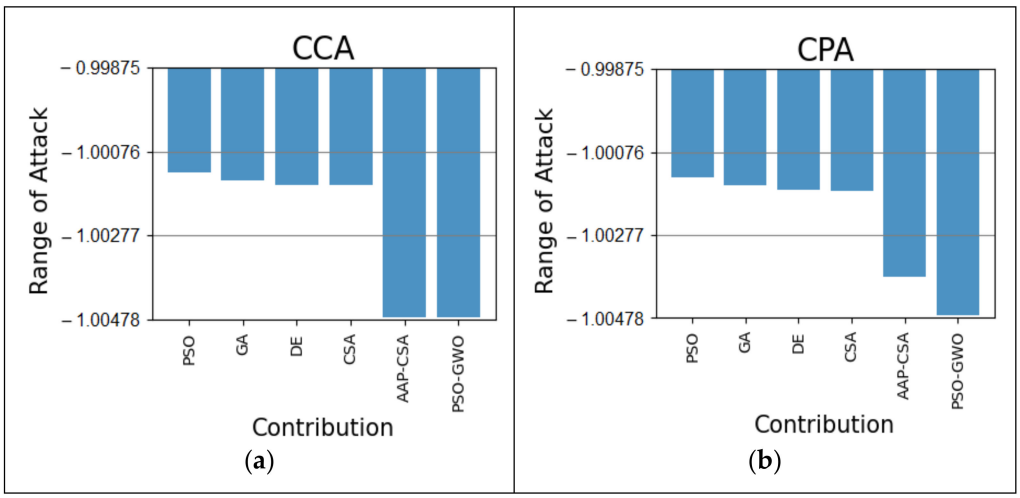

Figure 5. Analysis of the performance of various algorithms for the autism at 24 months dataset. (**a**) Performances based on the CCA attack; (**b**) Performances based on the CPA attack.

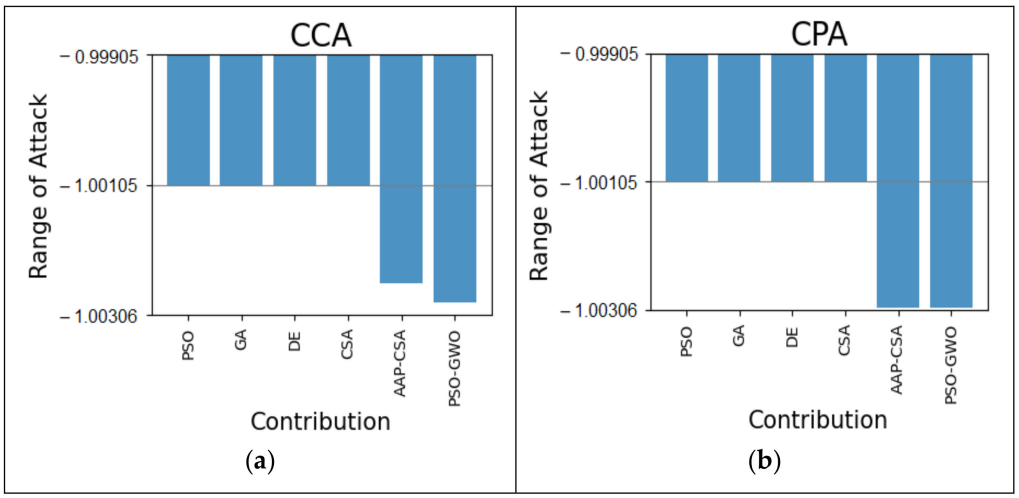

Figure 6. Analysis of the performance of various algorithms for the autism at 30 months dataset. (**a**) Performances based on the CCA attack; (**b**) Performances based on the CPA attack.

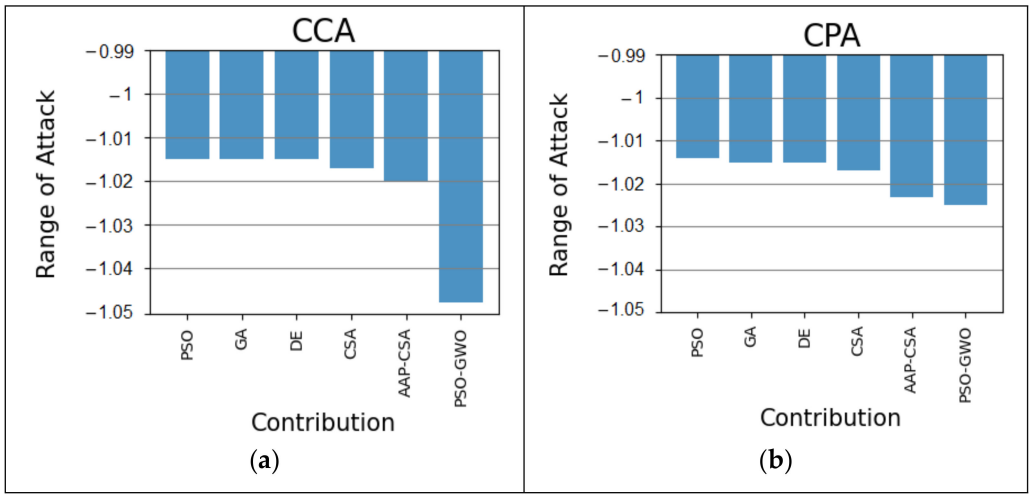

Figure 7. Analysis of the performance of various algorithms for the autism at 36 months dataset. (**a**) Performances based on the CCA attack; (**b**) Performances based on the CPA attack.

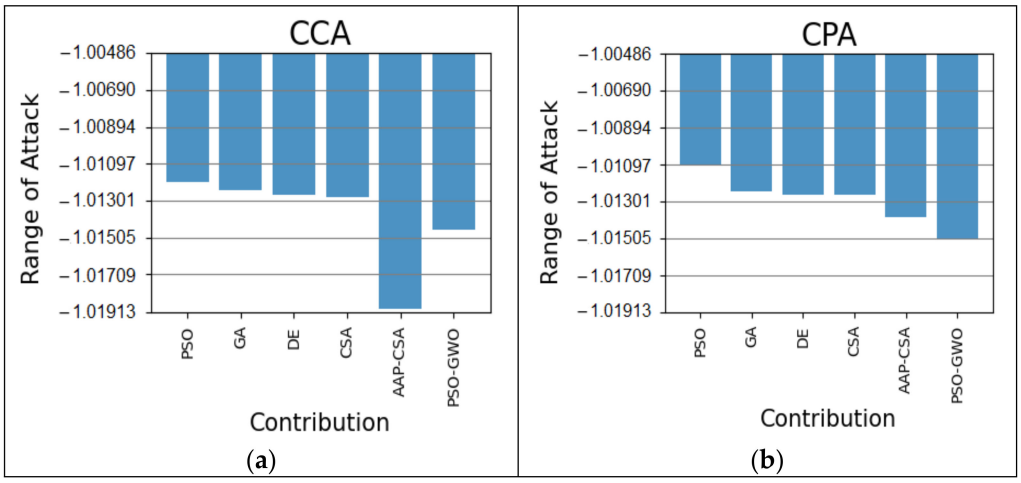

Figure 8. Analysis of the performance of various algorithms for the autism at 48 months dataset. (**a**) Performances based on the CCA attack; (**b**) Performances based on the CPA attack.

5. Results and Discussions

Among the different sorts of attacks, KCA and KPA were investigated initially and compared with other traditional algorithms revealed in Figure 4. From the simulation, the KCA attack over the proposed method is 0.44% superior to the PSO and GA and 0.43% more beneficial than DE and CSA in Figure 4a. Again, the KPA attack on the proposed method is 0.36 and 0.01% improved from PSO and AAP-CSA, as well as 0.31% enhanced in comparison with the remaining GA, DE, and CSA algorithms in Figure 4b. The overall results are shown in Table 3.

Table 3. The performance of enhanced combined PSO-GWO in terms of KCA and KPA attacks in comparison with the other algorithms.

PSO-GWO	PSO	GA	DE	CSA	AAP-CSA	Attacks
Superior to	0.44%	0.44%	0.43%	0.43%	x	KCA
Higher than	0.36%	0.31%	0.31%	0.31%	0.01%	KPA

On the other hand, our four types of autism datasets for the CCA and CPA attacks are demonstrated in Figures 5–8. Figure 5a shows that the proposed approach shows an

improvement of 0.37% and 0.33% compared to PSO and GA, respectively, and is 0.32% more beneficial than the DE and CSA algorithm, respectively, using the autism at 24 months dataset in terms of a CCA attack. For CPA analysis, our proposed scheme is 0.34%, 0.32%, 0.31%, 0.30%, and 0.10% more effective than the PSO, GA, DE, CSA, and AAP-CSA, respectively, as shown in Figure 5b. The total outcomes are summarized in Table 4.

Table 4. The performance of enhanced combined PSO-GWO in terms of CCA and CPA attacks in comparison with the other algorithms under the 24 months autism dataset.

PSO-GWO	PSO	GA	DE	CSA	AAP-CSA	Attacks
Enhanced over	0.37%	0.33%	0.32%	0.32%	x	CCA
Greater than	0.34%	0.32%	0.31%	0.30%	0.10%	CPA

For the autism at 30 months dataset, our method, in terms of the CCA attack, is 0.03% better than AAP-CSA and 0.18% superior to all other typical algorithms, as shown in Figure 6a. Similarly, the CPA attack is also 0.20% superior to PSO, GA, DE, and CSA, as shown in Figure 6b. Table 5 displays the results discussed above.

Table 5. The performance of enhanced combined PSO-GWO in terms of CCA and CPA attacks in comparison with the other algorithms for the autism at 30 months dataset.

PSO-GWO	PSO	GA	DE	CSA	AAP-CSA	Attacks
Superior to	0.18%	0.18%	0.18%	0.18%	0.03%	CCA
Higher than	0.20%	0.20%	0.20%	0.20%	x	CPA

The CCA attack on the autism at 36 months dataset is 3.30% better than PSO, GA, DE, 3.10% superior to CSA, and 2.80% better than AAP-CSA, which is illustrated in Figure 7a. In addition, our method for the CPA analysis on the 36 months autism dataset is 1.10% more improved than PSO, 1% better than GA and DE, 0.80% superior to CSA algorithms, and 0.20% better than AAP-CSA, as shown in Figure 7b. The performances are depicted in Table 6.

Table 6. The performance of enhanced combined PSO-GWO in terms of CCA and CPA attacks in comparison with the other algorithms for the autism at 36 months dataset.

PSO-GWO	PSO	GA	DE	CSA	AAP-CSA	Attacks
Excellent over	3.30%	3.30%	3.30%	3.10%	2.80%	CCA
Greater than	1.10%	1%	1%	0.80%	0.20%	CPA

In the case of the autism at 48 months dataset, the CCA simulation for our scheme is 0.26%, 0.23%, 0.22%, and 0.18% better than the PSO, GA, DE, and CSA algorithms, accordingly, as illustrated in Figure 8a. In Figure 8b, the CPA attack on the autism at 48 months dataset is 0.40% better than PSO, 0.29% superior to GA, 0.28% higher than DE and CSA, and 0.10% better than AAP-CSA. In this regard, the overall results are shown in Table 7.

Table 7. The performance of enhanced combined PSO-GWO in terms of CCA and CPA attacks in comparison with the other algorithms under the 48 months autism dataset.

PSO-GWO	PSO	GA	DE	CSA	AAP-CSA	Attacks
Better than	0.26%	0.23%	0.22%	0.18%	x	CCA
Superior to	0.40%	0.29%	0.28%	0.28%	0.10%	CPA

Thus, the simulation demonstrates that our proposed information security technique performed better than the existing conventional algorithms based on some attacks. There-

fore, it is revealed from the simulation outcomes that our sanitizing approach performs more effectively and efficiently compared to other existing traditional algorithms.

Due to the fact that sensitive diagnostic data of autism are critical for determining whether an individual is autistic or not, protecting this type of data is critical, which has greater applicability in the healthcare sector. Evidence produced by this study showed that our proposed sanitizing approach protects these data better than existing algorithms against certain attacks. It is, however, suggested that our recommended approach can be widely applied to the healthcare sector for data security and privacy.

6. Conclusions

The security and privacy of the autism dataset through the sanitizing technique were investigated in this study. The emphasis of this method was to conceal the sensitive data of patients. Specifically, an optimal key was produced for concealing the sensitive data, which was selected by the proposed Enhanced Combined PSO-GWO framework and resolved the problems mentioned in introduction. Furthermore, the results obtained by our recommended model were compared with existing traditional algorithms for justification. Mainly, our suggested technique was tested in terms of the different attacks and compared with existing traditional algorithms, and the expected outcomes were achieved, according to the experimental review. Our proposed technique, for the autism at 24 months dataset in terms of the CCA attack, is 0.37% and 0.33% better than the algorithms of PSO and GA, respectively, and 0.32% better than DE and CSA individually. Additionally, the suggested approach, in the case of the CPA attack, shows 0.20% more improvement compared to the PSO, GA, DE, and CSA algorithms, for the autism at 30 months dataset. For the autism at 36 months dataset, the simulation result of the proposed technique with CCA attacks is 3.30% more improved than PSO, GA, and DE, 3.10% better enhanced from CSA, and 2.80% superior to the AAP-CSA algorithms. Finally, in terms of the CPA attack on the autism at 48 months dataset, our technique is 0.40%, 0.29%, 0.28%, and 0.28% superior to PSO, GA, DE, and CSA, and 0.10% better than AAP-CSA, respectively.

Therefore, it is revealed from the analyzed results that our proposed enhanced technique is more effective and efficient compared with the present conventional algorithms.

In contrast, our future research will focus on improving a restoration technique in which the optimal key will be used in information security, specifically for the security and privacy of autism data.

Author Contributions: Conceptualization, M.M.R.; methodology, M.M.R. and R.C.M.; software, M.M.R. and S.S.; validation, M.M.R., R.C.M., S.S. and S.M.; formal analysis, M.M.R. and R.C.M.; investigation, M.M.R., R.C.M. and S.S.; resources, M.M.R., R.C.M., S.S. and S.M.; data curation, M.M.R. and S.M.; writing—original draft preparation, M.M.R.; writing—review and editing, M.M.R., R.C.M., S.S. and S.M.; visualization, M.M.R. and R.C.M.; supervision, R.C.M., S.S. and S.M.; project administration, R.C.M.; funding acquisition, R.C.M. and S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by Universiti Kebangsaan Malaysia (UKM) and the Department of Higher Education, Malaysia Education Ministry, grant nos. GGP-2019-023, GG-2019-059, and GPK/4IR/2020-022.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The Autism Child Datasets were collected from the Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia (accessed on 25 June 2020).

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

Abbreviations	**Explanations**
AAP-CSA	Adaptive Awareness Probability-based Crow Search Algorithm
ABC	Artificial Bee Colony
AES	Advanced Encryption Standard
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASD	Autism Spectrum Disorder
BF	Beamforming
BS-WOA	Brain Storm-based Whale Optimization
CCA	Chosen Cipher Attack
CI-LA	Crossover Improved-Lion Algorithm
CloudDLP	Cloud Data Loss Prevention
CPA	Chosen Plaintext Attack
CP-ABE	Ciphertext-Policy Attribute-Based Encryption
CSA	Crow Search Algorithm
CSA	Cuckoo Search Algorithm
DE	Differential Evolution
DES	Data Encryption Standard
DLP	Data Loss Prevention
DM	Degree of Modification
DSM-5	Diagnostic and Statistical Manual of Mental Disorders 5th edition
EPR	Electronic Patient Record
FR	False Rule
GA	Genetic Algorithm
GMGW	Genetically Modified Glowworm Swarm Optimization
GSA	Gravitational Search Algorithm
GSO	Glowworm Swarm Optimization
GWO	Grey Wolf Optimization
HF	Hiding Failure
IDEA	International Data Encryption Algorithm
IEP	Individual Education Plan
IFSP	Individualized Family Service Program
IoT	Internet of Things
IP	Information Preservation
JA	Jaya Algorithm
J-SSO	Jaya-based Shark Smell Optimization
KCA	Known Cipher Attack
KPA	Known Plaintext Attack
LSB	Least Significant Bit
OI-CSA	Opposition Intensity-based Cuckoo Search Algorithm
PASH	Privacy-Aware Smart Health
PGVIR	Parallelized Grouped Victim Item Removal
PHCR	Parallelized Hiding Candidate Removal
PSO	Particle Swarm Optimization
PSO-GWO	Particle Swarm Optimization- Grey Wolf Optimization
RC4	Rivest Cipher 4
ROI	Region Of Interest
RONI	Region Of Non-Interest
RSMA	Rate-Splitting Multiple Access
SAIN	Satellite and Aerial-Integrated Network
SHRs	S-Health Records
SSO	Shark Smell Optimization
TripleDES	Triple Data Encryption Standard
UAV	Unmanned Aerial Vehicle

UKM	Universiti Kebangsaan Malaysia
WNU	Whale with New Crosspoint-based Update
WRS	Wilcoxon Rank Sum

List of Mathematical Symbols

Symbols	Descriptions
D	Processed (from original) database
D'	Sanitization database
$K_1, K_2, \ldots K_N$	Number of keys
K_2	Pruned key matrix
$\oplus$	XOR operator
$+$	Binary Summation
$\lfloor \rfloor$	Floor function
L_{D}	Sanitization key length
$\sqrt{L \frac{C}{D}}$	Key length
$\mathrm{T}_1, \mathrm{T}_2, \ldots \mathrm{T}_5$	Number of transactions
$\mathrm{T}_{\max}$	Maximum transaction
$\otimes$	Kronecker product
C_1, C_2, C_3	Objective functions
f_{s}	Frequency of sensitive itemset in sanitized data
f_m	Frequency of sensitive itemset in original data
f_{ns}	Frequency of non-sensitive itemset in sanitized data
$w1, w2, w3$	Impact of a particular cost function
f	Fitness function
G	Minimum objective function
$\vec{M}$	Location of the particle
$\vec{w}$	Velocity of the particle
ω	User-defined behavioral parameter (an inertia weight)
$\vec{q}$	Particle's previous best position (pbest position)
$\vec{f}$	Particle's previous best position in the swarm (gbest position)
r_1, r_2	Stochastic variables
c_1, c_2	Acceleration constants
u	Current iteration
$\vec{H}, \vec{E}$	Coefficient vectors

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders DSM-5®*, 5th ed.; American Psychiatric Association: Arlington, VA, USA, 2013.
2. Thabtah, F. Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. In Proceedings of the 1st International Conference on Medical and health Informatics, Taichung City, Taiwan, 20–22 May 2017; pp. 1–6. [CrossRef]
3. Sahmim, S.; Gharsellaoui, H. Privacy and Security in Internet-based Computing: Cloud Computing, Internet of Things, Cloud of Things: A review. *Procedia Comput. Sci.* **2017**, *112*, 1516–1522. [CrossRef]
4. Sivan, R.; Zukarnain, Z.A. Security and Privacy in Cloud-Based E-Health System. *Symmetry* **2021**, *13*, 742. [CrossRef]
5. Mewada, S.; Gautam, S.S.; Sharma, P. Artificial Bee Colony-Based Approach for Privacy Preservation of Medical Data. *Int. J. Inf. Syst. Modeling Des.* **2020**, *11*, 22–39. [CrossRef]
6. Mandala, J.; Rao, M.V.P.C.S. Privacy preservation of data using crow search with adaptive awareness probability. *J. Inf. Secur. Appl.* **2019**, *44*, 157–169. [CrossRef]
7. Rahman, M.M.; Usman, O.L.; Muniyandi, R.C.; Sahran, S.; Mohamed, S.; Razak, R.A. A Review of Machine Learning Methods of Feature Selection and Classification for Autism Spectrum Disorder. *Brain Sci.* **2020**, *10*, 949. [CrossRef] [PubMed]
8. The Lancet Neurology. Investing in autism: Better evidence for better care. *Lancet Neurol.* **2017**, *16*, 251. [CrossRef]
9. Fein, D.; Helt, M. Facilitating Autism Research. *J. Int. Neuropsychol. Soc.* **2017**, *23*, 903–915. [CrossRef] [PubMed]
10. Volkmar, F.R.; Reichow, B.; McPartland, J. Classification of autism and related conditions: Progress, challenges, and opportunities. *Dialogues Clin. Neurosci.* **2012**, *14*, 229–237. [CrossRef]
11. Seh, A.H.; Zarour, M.; Alenezi, M.; Sarkar, A.K.; Agrawal, A.; Kumar, R.; Khan, R.A. Healthcare Data Breaches: Insights and Implications. *Healthcare* **2020**, *8*, 133. [CrossRef]

12. Vezyridis, P.; Timmons, S. Resisting big data exploitations in public healthcare: Free riding or distributive justice? *Sociol. Health Illn.* **2019**, *41*, 1585–1599. [CrossRef]
13. Coventry, L.; Branley, D. Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *Maturitas* **2018**, *113*, 48–52. [CrossRef] [PubMed]
14. Price, W.N.; Cohen, I.G. Privacy in the age of medical big data. *Nat. Med.* **2019**, *25*, 37–43. [CrossRef] [PubMed]
15. Chernyshev, M.; Zeadally, S.; Baig, Z. Healthcare Data Breaches: Implications for Digital Forensic Readiness. *J. Med. Syst.* **2019**, *43*, 1–12. [CrossRef] [PubMed]
16. Raspa, M.; Moultrie, R.; Wagner, L.; Edwards, A.; Andrews, S.; Frisch, M.K.; Turner-brown, L.; Wheeler, A. Ethical, Legal, and Social Issues Related to the Inclusion of Individuals With Intellectual Disabilities in Electronic Health Record Research: Scoping Review. *J. Med. Internet Res.* **2020**, *22*, e16734. [CrossRef] [PubMed]
17. Trustwave. *The Value of Data: A Cheap Commodity or a Priceless Asset?* Final Report; UK, 2017; Unpublished work.
18. Sulleyman, A. NHS cyber attack: Why stolen medical information is so much more valuable than financial data. *The Independent*. 12 May 2017. Available online: https://www.independent.co.uk/life-style/gadgets-and-tech/news/nhs-cyber-attack-medical-data-records-stolen-why-so-valuable-sell-financial-a7733171.html (accessed on 10 June 2021).
19. Kangas, E. Why Are Hackers Targeting Your Medical Records? 2017. Available online: https://luxsci.com/blog/hackers-targeting-medical-records.html (accessed on 10 June 2021).
20. Şenel, F.A.; Gökçe, F.; Yüksel, A.S.; Yiğit, T. A novel hybrid PSO–GWO algorithm for optimization problems. *Eng. Comput.* **2018**, *35*, 1359–1373. [CrossRef]
21. Zolghadr-Asli, B.; Bozorg-Haddad, O.; Chu, X. Crow Search Algorithm (CSA). In *Advanced Optimization by Nature-Inspired Algorithms. Studies in Computational Intelligence*; Bozorg-Haddad, O., Ed.; Springer: Singapore, 2018; Volume 720. [CrossRef]
22. Alphonsa, M.M.A.; Amudhavalli, P. Genetically modified glowworm swarm optimization based privacy preservation in cloud computing for healthcare sector. *Evol. Intell.* **2018**, *11*, 101–116. [CrossRef]
23. Panda, M.; Das, B. Grey Wolf Optimizer and Its Applications: A Survey. In Proceedings of the Third International Conference on Microelectronics, Computing and Communication Systems, ARTTC BSNL, Ranchi, India, 12–13 May 2018; Springer: Singapore; pp. 179–194. [CrossRef]
24. Ahamad, D.; Hameed, S.A.; Akhtar, M. A multi-objective privacy preservation model for cloud security using hybrid Jaya-based shark smell optimization. *J. King Saud. Univ.Comput. Inf. Sci.* **2020**. [CrossRef]
25. Balashunmugaraja, B.; Ganeshbabu, T.R. Optimal Key Generation for Data Sanitization and Restoration of Cloud Data: Future of Financial Cyber Security. *Int. J. Inf. Technol. Decis. Mak.* **2020**, *19*, 987–1013. [CrossRef]
26. Abidi, M.H.; Alkhalefah, H.; Umer, U.; Mohammed, M.K. Blockchain-based secure information sharing for supply chain management: Optimization assisted data sanitization process. *Int. J. Intell. Syst.* **2021**, *36*, 260–290. [CrossRef]
27. Edgar, D. Data Sanitization Techniques. A Net 2000 Ltd. White Paper, 2003–2004. Available online: http://www.orafaq.com/papers/data_sanitization.pdf (accessed on 20 June 2021).
28. Lekshmy, P.L.; Rahiman, M.A. A sanitization approach for privacy preserving data mining on social distributed environment. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *11*, 2761–2777. [CrossRef]
29. Ochôa, I.S.; Silva, L.A.; De Mello, G.; Garcia, N.M.; De Santana, J.F.P.; Leithardt, V.R.Q. A Cost Analysis of Implementing a Blockchain Architecture in a Smart Grid Scenario Using Sidechains. *Sensors* **2020**, *20*, 843. [CrossRef] [PubMed]
30. Shailaja, G.K.; Rao, C.V.G. Opposition Intensity-Based Cuckoo Search Algorithm for Data Privacy Preservation. *J. Intell. Syst.* **2019**, *29*, 1441–1452. [CrossRef]
31. Renuga, S.; Jagatheeshwari, S.S.K. Efficient Privacy-Preserving Data Sanitization over Cloud Using Optimal GSA Algorithm. *Comput. J.* **2018**, *61*, 1577–1588. [CrossRef]
32. Han, P.; Liu, C.; Cao, J.; Duan, S.; Pan, H.; Cao, Z.; Fang, B. CloudDLP: Transparent and Scalable Data Sanitization for Browser-Based Cloud Storage. *IEEE Access* **2020**, *8*, 68449–68459. [CrossRef]
33. Revathi, S.T.; Ramaraj, N.; Chithra, S. Brain storm-based Whale Optimization Algorithm for privacy-protected data publishing in cloud computing. *Clust. Comput.* **2018**, *22*, 3521–3530. [CrossRef]
34. Sahi, A.; Lai, D.; Li, Y. Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan. *Comput. Biol. Med.* **2016**, *78*, 1–8. [CrossRef] [PubMed]
35. Liu, Y.; Qu, X.; Xin, G. A ROI-based reversible data hiding scheme in encrypted medical images. *J. Vis. Commun. Image Represent.* **2016**, *39*, 51–57. [CrossRef]
36. Zhang, Y.; Zheng, D.; Deng, R.H. Security and Privacy in Smart Health: Efficient Policy-Hiding Attribute-Based Access Control. *IEEE Internet Things J.* **2018**, *3*, 1–15. [CrossRef]
37. Sharma, U.; Toshniwal, D.; Sharma, S. A sanitization approach for big data with improved data utility. *Appl. Intell.* **2020**, *50*, 2025–2039. [CrossRef]
38. Sharma, S.; Toshniwal, D. MR-OVnTSA: A heuristics based sensitive pattern hiding approach for big data. *Appl. Intell.* **2020**, *50*, 4241–4260. [CrossRef]
39. Lin, Z.; Lin, M.; Champagne, B.; Zhu, W.-P.; Al-Dhahir, N. Secure and Energy Efficient Transmission for RSMA-Based Cognitive Satellite-Terrestrial Networks. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 251–255. [CrossRef]
40. Lin, Z.; Lin, M.; De Cola, T.; Wang, J.-B.; Zhu, W.-P.; Cheng, J. Supporting IoT With Rate-Splitting Multiple Access in Satellite and Aerial-Integrated Networks. *IEEE Internet Things J.* **2021**, *8*, 11123–11134. [CrossRef]

41. Lin, Z.; Lin, M.; Champagne, B.; Zhu, W.-P.; Al-Dhahir, N. Secrecy-Energy Efficient Hybrid Beamforming for Satellite-Terrestrial Integrated Networks. *IEEE Trans. Commun.* **2021**, *69*, 6345–6360. [CrossRef]
42. Mottalib, M.M.; Rahman, M.M.; Habib, M.T.; Ahmed, F. Detection of the Onset of Diabetes Mellitus by Bayesian Classifier Based Medical Expert System. *Trans. Mach. Learn. Artif. Intell.* **2016**, *4*, 1–8. [CrossRef]
43. Rahman, M.A.; Muniyandi, R.C.; Albashish, D.; Rahman, M.M.; Usman, O.L. Artificial neural network with Taguchi method for robust classification model to improve classification accuracy of breast cancer. *PeerJ Comput. Sci.* **2021**, *7*, 1–27. [CrossRef]
44. Rahman, M.A.; Muniyandi, R.C.; Islam, K.T.; Rahman, M.M. Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model. In Proceedings of the 2019 IEEE Student Conference on Research and Development (SCOReD), Perak, Malaysia, 15–17 October 2019; IEEE: Perak, Malaysia, 2019; pp. 33–38. [CrossRef]
45. Alesawy, O.; Muniyandi, R.C. Elliptic Curve Diffie-Hellman Random Keys Using Artificial Neural Network and Genetic Algorithm for Secure Data over Private Cloud. *Inf. Technol. J.* **2016**, *15*, 77–83. [CrossRef]
46. Usman, O.L.; Muniyandi, R.C. CryptoDL: Predicting Dyslexia Biomarkers from Encrypted Neuroimaging Dataset Using Energy-Effcient Residue Number System and Deep Convolutional Neural Network. *Symmetry* **2020**, *12*, 836. [CrossRef]
47. Bostanoğlu, B.E.; Öztürk, A.C. Minimizing information loss in shared data: Hiding frequent patterns with multiple sensitive support thresholds. *Stat. Anal. Data Min. ASA Data Sci. J.* **2020**, *13*, 309–323. [CrossRef]
48. Iwendi, C.; Moqurrab, S.A.; Anjum, A.; Khan, S.; Mohan, S.; Srivastava, G. N-Sanitization: A semantic privacy-preserving framework for unstructured medical datasets. *Comput. Commun.* **2020**, *161*, 160–171. [CrossRef]
49. Zaman, A.N.K.; Obimbo, C.; Dara, R.A. An Improved Data Sanitization Algorithm for Privacy Preserving Medical Data Publishing. In *Advances in Artificial Intelligence, Proceedings of the Canadian AI 2017 Lecture Notes in Computer Science, Edmonton, AB, Canada, 16–19 May 2017*; Mouhoub, M., Langlais, P., Eds.; Springer: Cham, Switzerland, 2017; pp. 64–70. [CrossRef]
50. Liu, X.; Chen, G.; Wen, S.; Song, G. An Improved Sanitization Algorithm in Privacy-Preserving Utility Mining. *Math. Probl. Eng.* **2020**, *2020*, 1–14. [CrossRef]
51. Freitas, W., Jr.; Favier, G.; De Almeida, A.L.F. Generalized Khatri-Rao and Kronecker Space-Time Coding for MIMO Relay Systems with Closed-Form Semi-Blind Receivers. *Signal* **2018**, *151*, 19–31. [CrossRef]