symmetry

MDPI

Article

PA-FRIM: An Adaptive Hybrid FOX–RUN Framework with Adaptive Intensive Mutation for Multi-Metric Big Data Anonymization

M. Faruk Şahin [1,*] and Can Eyüpoğlu [2]

[1] Department of Computer Engineering, Istanbul Atlas University, 34408 Istanbul, Türkiye
[2] Department of Computer Engineering, Turkish Air Force Academy, National Defence University, 34149 Istanbul, Türkiye; can.eyupoglu@msu.edu.tr
* Correspondence: muhammed.sahin@atlas.edu.tr

Abstract

Background/Objectives: Privacy preservation in big data environments is an NP-hard optimization task that requires the satisfaction of k-anonymity and l-diversity constraints to ensure data utility. **Methods:** This study proposes a novel hybrid optimization approach, adaptive hybrid FOX–RUN Intensive Mutation (PA-FRIM), to address the privacy–utility trade-off in anonymization process. The proposed approach integrates FOX-based global exploration with RUN-based local search using a hybrid adaptive control strategy and intensive mutation search to improve solution diversity in highly constrained solution spaces. **Results:** The experimental study on the Adult and Bank Marketing datasets shows that PA-FRIM exhibits stable convergence behavior compared to competing methods. The results indicate that full privacy is achieved on the Adult dataset with a violation value of 0.00, and information loss is minimized with an NIL measure of 0.5686. From the analytical utility perspective, PA-FRIM ensures data usability, even in the constrained region, achieving classification accuracies of 89.61% on the Bank Marketing dataset and 84.90% on the Adult dataset. **Conclusions:** By using a multi-metric evaluation strategy, PA-FRIM provides a robust optimization framework that eliminates privacy violations while maintaining high analytical performance.

Keywords: privacy-preserving data publishing; big data anonymization; k-anonymity; l-diversity; hybrid metaheuristic optimization

Check for updates

Academic Editor: Shi Cheng

Received: 5 March 2026
Revised: 4 April 2026
Accepted: 23 April 2026
Published: 25 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In recent years, large datasets emerging from fields such as healthcare [1], finance [2], social media [3], and various other domains offer significant opportunities in data analytics processes due to their high volume, heterogeneous structures, and complex relationships [4]. However, the protection of sensitive information within these vast datasets has become a central necessity within the framework of global legal regulations and ethical standards [5,6]. Privacy preservation in big data environments is commonly achieved through structural models such as k-anonymity, l-diversity, and t-closeness in the literature [7,8]. However, these traditional approaches fall short in managing the natural trade-off between the analytical utility of data and the level of privacy in high-dimensional and discrete search spaces [9–11]. Specifically, as the volume and velocity of big data increase, determining the ideal generalization levels that minimize information loss becomes an NP-hard problem. This results in the anonymization process evolving from a simple constraint satisfaction

problem to an optimization problem involving multi-metric and conflicting objectives. At this point, the computational cost and performance bottlenecks of deterministic methods make use of metaheuristic algorithms, which offer the ability to converge toward the global optimum in complex search spaces, a necessity [12,13].

Metaheuristic algorithms are notable for their ability to balance exploration and exploitation in complex optimization problems [14]. Furthermore, metaheuristic-based optimization methods focus on multi-objective problems in complex spaces such as big data, but these problems are typically static and assumed to be invariant throughout the optimization process [15]. This static structure prevents the algorithm from adapting to the complex constraint regions and instantaneous convergence dynamics that result from the heterogeneous nature of big data. Because of the complexity of the anonymization problem and the limited number of valid solution regions, algorithms become stuck in local optima [16,17]. In particular, when strict privacy constraints such as k-anonymity and l-diversity are combined, single-phase and static search strategies are insufficient to avoid becoming stuck in such narrow solution regions. In this context, the absence of hybrid metaheuristic studies with hybrid adaptive mechanisms capable of adapting to the dynamic nature of big data and providing more effective search and perturbation strategies represents a significant methodological gap in the literature. Metaheuristic studies in the existing literature generally address the anonymization problem through singular algorithmic structures or static hybrid approaches. However, these approaches encounter a methodological bottleneck in managing the rugged and fragmented constraint spaces induced by the simultaneous enforcement of the k-anonymity and l-diversity models. Theoretically, while k-anonymity renders the search space discrete by enforcing that records must exist in similarity groups of at least size k, l-diversity further constrains the feasible solution regions by regulating the distribution of sensitive attributes. This situation shapes the performance limits of individual algorithms within the framework of the No Free Lunch theorem, leading to a structural imbalance between global exploration across the expansive search space and local exploitation within the narrowly constrained feasible regions. Therefore, the technical problem is not merely the application of a single metaheuristic method, but rather the absence of a phase-adaptive mechanism capable of overcoming local optima traps within these complex and constrained topologies.

Motivation and Contribution

In recent years, the sharing of large datasets for analysis has demanded a subtle trade-off between privacy preservation and data utility. The anonymization methods in the current literature are mainly rooted in one-dimensional optimization approaches. Moreover, there are difficulties in reaching the minimum information loss while satisfying the privacy constraints and ensuring the dataset is still amenable to analysis. Especially in high-dimensional datasets with sparse feature distributions, it is difficult to search for solutions satisfying constraints such as k-anonymity and l-diversity simultaneously. This leads to the fact that the classical meta-heuristic algorithms are often faced with severe constraints in the search process, result in population diversity loss, get trapped in local optima, and tend to have overgeneralization, with strong effects on data usability. In this regard, the paper proposes an optimization procedure for addressing the trade-off between privacy constraints and data utility in the anonymization problem, which is sensitive to multiple criteria and constraints. A hybrid adaptive FOX-RUN Intensive Mutation (PA-FRIM) algorithm is proposed for this optimization procedure, which can maintain the balance between exploration and exploration in the limited solution space through phase adaptation and differential refinement. The proposed hybrid adaptive algorithm is expected to achieve a balance between optimization in complex and limited search spaces

that occur during the anonymization process by integrating the global exploration ability of the FOX algorithm with the differential refinement of the Runge–Kutta algorithm in a hybrid adaptive framework. Moreover, the adaptive intensive mutation phase of the proposed algorithm is expected to alleviate the risk of early convergence by turning on during the exploitation phase and maintaining population diversity in the narrow regions defined by the privacy constraints. The principal distinction and originality of the PA-FRIM algorithm, relative to existing hybrid metaheuristic frameworks in the literature, stem from its design around a phase-adaptive knowledge transfer mechanism that extends beyond a static coupling between exploration and exploitation phases. Whereas conventional hybrid models typically employ two algorithms either sequentially or concurrently, the proposed approach filters high-quality structural information obtained during the FOX phase using Cauchy-based jump operators and dynamically injects it—via elite individuals—into the differential refinement process of the RUN phase. This inter-phase adaptive coordination enables the algorithm not merely to perform search, but also to overcome the steep gradients induced by anonymization constraints in a manner akin to a learning process. Moreover, the Adaptive Intensive Mutation layer activated during the exploitation phase operates in coordination with the mathematical precision of the RUN algorithm, actively eliminating local optima traps by stochastically evaluating the positions of candidate solutions within the generalization hierarchy through a feedback-driven mechanism. This integrated architecture elevates PA-FRIM beyond a simple combination of two algorithms, transforming it into a novel hybrid management framework tailored for constraint-aware optimization processes.

In this respect, the main contributions of the study to the existing literature are:

- The data anonymization problem is defined as a multi-metric and constraint-sensitive optimization problem that focuses on the interaction between privacy (k-anonymity and l-diversity) and data utility.
- The PA-FRIM algorithm, which combines FOX and RUN-based search strategies under a hybrid adaptive structure and is supported by differential updates, is proposed.
- The optimization success is comprehensively analyzed using the Normalized Information Loss (NIL), Discernibility Metric (DM), Average Equivalence Class Size (C_{avg}), Classification Accuracy (ACC), k-Anonymity Violation (k_v), and l_V metrics, which directly reflect anonymization quality.
- Through experiments conducted on real-world datasets with different statistical distributions, the proposed approach is validated for its privacy compliance and competitive performance.

2. Related Works

2.1. Traditional and Contemporary Heuristic Methods for Big Data Privacy

The dynamic nature of big data ecosystems necessitates a delicate balance between privacy protection and data utility in anonymization procedures. In the existing literature, methods for balancing trade-offs are modeled using deterministic and heuristic models. In this context, a novel chaos and corruption algorithm is introduced to counter attribution attacks that arise during the sharing of big data [8]. The proposed algorithm is validated using Kullback–Leibler bias and classification accuracy measures, which leverage the stochastic process of chaos theory. The accuracy values for Voted Perceptron, OneR, Naive Bayes, and Decision Tree classifiers are 78.42, 80.22, 82.64, and 85.69, respectively. The proposed chaos model performs better than conventional studies in terms of Kullback–Leibler bias and statistical efficiency for F-measure values. However, the sensitivity of the chaos parameters, as well as the computational complexity for large datasets, limit the algorithm's adaptability to various data architectures. Outliers in the data are another

significant aspect that directly impact data utility. To address this issue, a hybrid method is proposed that combines the connectivity-based outlier factor (COF) with the conventional Mondrian algorithm [11]. In the proposed algorithm, the COF process detects the outliers, which reduces the impact of data on the generalization process. The experimental results reveal that it performs better than the conventional Mondrian algorithm in terms of the discernibility metric (DM), normalized average equivalence class size, and global certainty penalty. However, it lacks an adequate analysis of the model's theoretical robustness against privacy attacks, which is a disadvantage. Another challenge in big data is the increased privacy risks associated with the continuous release of data in dynamic data streams. To address this issue, the Improved Distributed M-Invariance (ImDMI) method is proposed, which implements the m-invariance privacy model using a distributed bottom-up R-tree approach on Apache Spark [18]. The proposed method is more efficient than existing techniques in terms of execution time and fake data generation, as it meets the m-invariance requirements during continuous data addition and separation processes. On the other hand, another study aiming to enhance the usability of k-anonymity models proposes a new information loss function based on information quantity theory and an improved 2-means clustering algorithm [19]. The proposed method minimizes information loss by assigning different weights to semi-qualitative attributes and improves stability in the selection of cluster centers using the average center method. However, these heuristic weighting and clustering methods struggle to overcome local optimum traps in high-dimensional data. In recent years, hybrid frameworks have been proposed to address the limitations of individual privacy models. In this regard, a novel r, k, ϵ-anonymization algorithm is introduced, which integrates the k-anonymity and ϵ-differential privacy models [7]. The proposed algorithm is developed following the General Data Protection Regulation (GDPR) guidelines. Experiments performed on real-world datasets indicate that the proposed algorithm yields a 31.74% greater increase in data utility than the k-anonymity model and a 26.99% increase than the ϵ-differential privacy model. The 30% improvement in the average error rate indicates the viability of hybrid models. Nevertheless, all these heuristic and rule-based strategies explored in this section are incapable of ensuring convergence to the global optimum in broad and discrete search spaces because of the NP-hardness of the anonymization problem. This gap emphasizes the need for more adaptable meta-heuristic optimization frameworks that avoid entrapment in complex environments.

2.2. Metaheuristic Optimization for Privacy-Preserving

Considering the dynamic nature of big data systems and environments, it is necessary to apply metaheuristic optimization techniques to reach the global optimum solution in complex search spaces. In this case, in Internet of Things (IoT) environments, in which high-dimensional data are generated, it is necessary to maintain the privacy of the data and the efficiency of analysis. To address this issue in high-dimensional big data environments, the PPSLOA-HDBDE method is proposed [20]. The proposed method minimizes the dimensionality problem of big data systems through feature selection using the Cat Swarm Optimization (SCSO) method and parameter tuning using the Marine Predators Algorithm (IMPA), reaching an accuracy level of 99.49%. However, in IoT systems in which the network is dynamically changing, the application of server-based systems may not be efficient in preventing collaborative attacks. In this case, the key generation method using Grasshopper–Black Hole Optimization (G-BHO) in Industrial IoT (IIoT) systems is proposed [21]. The proposed method defines a multi-objective function that includes parameters such as the degree of modification, hiding rate, correlation coefficient between original and restored data, and information preservation rate, and uses this function to generate the optimal key. From the perspective of privacy protection, the proposed G-BHO

outperforms JA, GWO, GOA, and BHO by 1%, 15.2%, 12.6%, and 1%, respectively. This ensures better adaptability to the complex dynamics of IoT environments. In another study, a technique for optimal key generation based on MPWO is proposed [22]. The proposed algorithm integrates the hunting behavior of marine predators with the characteristics of whales to determine the optimal key. The optimal key generated using the MPWO algorithm effectively protects data privacy, with the following results: disclosure attack: 0.03, identity disclosure attack: 185.07, and no attack: 0.04. Similarly, the discernibility metrics for disclosure attack, identity disclosure attack, and no attack are 0.08, 123.38, and 0.09, respectively, while the Normalized Certainty Penalty (NCP) is 0.002, 61.69, and 0.001, respectively. The integration of these sanitization techniques into cloud-based Electronic Health Records (EHR) systems is diversified using the Seagull Optimization Algorithm (SOA) [23]. Random key generation based on SOA provides a security level of 98.58%. However, while such key-based encryption methods ensure static data security, they remain limited in minimizing structural information loss (utility loss) during analytical data sharing. To overcome the NP-hard nature of clustering-based approaches in k-anonymity models that structurally optimize both data privacy and utility, a Black Hole Algorithm (BHA)-based method is proposed [24]. The proposed approach improves data utility compared to k-anonymity, BHA-based k-anonymity, and clustering-based k-anonymity approaches. The optimization success provided by BHA is further enhanced through a more dynamic structure proposed using the Adaptive Whale Optimization Algorithm (AWOA) for social network data [25]. The proposed method uses a swarm-based clustering technique to group social data. AWOA minimizes information loss by 11% for the Yelp dataset and 10% for the Facebook dataset. However, its use on large datasets with a large number of attributes is not practical. To address this limitation, a multilayered framework is proposed called Distributed Privacy Preservation (DPP) for OSN using Artificial Intelligence (DPP-OSN-AI) [26]. It is based on the Whale Optimization Algorithm. In the proposed framework, the k-anonymity, l-diversity, and t-proximity models are applied in a systematic hierarchy to create a structure that is more resistant to data breaches. The experimental results show that this AI-based clustering strategy improves Degree of Anonymization (DoA) by 15.57% and Information Loss Efficiency (ILE) by 17.85%. However, this method is tested on static datasets and has limitations in managing dynamic parameters such as concept drift and data redundancy in real-time data streams. Moreover, the unexplored risks related to artificial intelligence pose a disadvantage for the model's operational robustness. To improve the scalability and complexity issues encountered in large-scale practical applications and high-dimensional attribute spaces, a Genetic Algorithm-based KGen method is proposed to calculate a permutation of the dataset that is both k-anonymized and usable for further processing [27]. The proposed method enables data managers to maintain high levels of anonymity while preserving usability and preventing information entropy loss on the data. However, most of the metaheuristic models examined here are based on single-phase evolutionary processes, which pose a risk of getting trapped in local optimum plateaus in high-dimensional privacy spaces. Although metaheuristic approaches examined in the existing literature have demonstrated notable success in managing the privacy–utility trade-off in the anonymization process, three fundamental methodological deficiencies remain, which the proposed PA-FRIM framework explicitly addresses. First, a substantial portion of prior studies relies on single-phase evolutionary processes, thereby facing a heightened risk of being trapped in local optima within the narrow feasible regions induced by k-anonymity and l-diversity constraints. Second, the absence of effective inter-phase knowledge transfer mechanisms in hybrid models limits the overall efficiency of the search process. Third, static search strategies fail to provide adaptive responses to topological variations in high-dimensional data spaces. In this context, the proposed

PA-FRIM architecture enhances the global exploration capability of the FOX algorithm through Cauchy-based jumps, enabling escape from tightly constrained feasible regions. Concurrently, it transfers high-quality structural information obtained during this phase to the RUN phase via elite-based knowledge transfer, where it is leveraged through differential refinement with high precision. Accordingly, unlike the static and monolithic approaches prevalent in the literature, PA-FRIM establishes a phase-adaptive hierarchy that addresses both the NP-hard nature of the anonymization problem and its constrained search topology through an integrated solution strategy.

3. Methodology

This section discusses the proposed PA-FRIM algorithm aimed at improving the steep constraint spaces and high-dimensional search topologies inherent in data anonymization problems. In this context, the objective function and penalty mechanism, which represent the balance between privacy constraints and data utility, are first defined. Subsequently, the functioning of the Cauchy-based jump and intensive mutation operators, supporting these processes through hybrid adaptive coordination of global exploration and local exploitation strategies, is discussed. Finally, the asymptotic computational complexity of the proposed hybrid structure is analyzed.

3.1. Definition of Objective Function

The data anonymization process is a multi-objective optimization problem that governs the trade-off between meeting privacy requirements and maximizing data utility. In this study, privacy constraints and data usage performance are combined into a single scalar fitness function. The objective function is structured to minimize structural information loss while bounding privacy violations through a quadratic penalty mechanism. The total fitness value is calculated according to Equation (1):

$$\mathrm{f}(\mathcal{X}) = \Phi(\mathcal{X}) + \Psi(\mathcal{X}) \tag{1}$$

Here, $\mathcal{X}$ represents the solution vector containing the generalization levels obtained from the anonymization process. The first component, $\Phi(\mathcal{X})$, is the penalty term that monitors violations of k-anonymity and l-diversity, as defined in Equation (2):

$$\Phi(\mathcal{X}) = \Lambda \cdot \left(V_k(\mathcal{X})^2 + V_\ell(\mathcal{X})^2\right) \tag{2}$$

Here, $\mathcal{V}_k(\mathcal{X})$ and $\mathcal{V}_\ell(\mathcal{X})$ represent the number of records and groups, respectively, that fall below the k and l threshold values. Multiplying these violations by the penalty coefficient Λ rapidly increases the fitness value of infeasible solutions, thus driving the algorithm towards regions where the constraints are satisfied. When the privacy constraints are satisfied, the optimization process then focuses on the data utility component, as defined in Equation (3):

$$\Psi(\mathcal{X}) = \alpha \cdot \mathrm{NIL}(\mathcal{X}) + \beta \cdot \mathrm{DM}(\mathcal{X}) \tag{3}$$

Here, Normalized Information Loss (NIL) and Discernibility Metric (DM) are cost elements that measure the reduction in the analytical value of the data. The weight coefficients α and β determine the priority balance between information loss and the discernibility of records. The NIL metric penalizes the generalization level, while the DM metric penalizes record clustering within equivalence classes. The mathematical characteristics of the proposed objective function are analyzed across three fundamental dimensions, in accordance with the inherent complexity of the anonymization problem. First, the function exhibits a discontinuous and non-convex structure. This arises from

the fact that each step within the generalization hierarchy induces abrupt and step-wise information loss in the data, while violations of the k and l constraints result in sharp jumps in the function value due to the imposed penalty coefficient. Second, the linear combination of the NIL and DM metrics is adopted because both measures are normalized within the range of [0, 1], and because they represent distinct dimensions of data utility (i.e., generalization depth versus record loss) that must be equally weighted. This linear formulation ensures the preservation of weight balance during the optimization process while simultaneously minimizing computational overhead. Third, the high value assigned to the penalty coefficient Λ (e.g., $\Lambda = 10^5$) enforces a death penalty strategy within the algorithm. This theoretical design compels the algorithm to immediately discard any infeasible solution—regardless of its utility—thereby transforming infeasible regions of the search space into insurmountable steep barriers and constraining PA-FRIM to operate exclusively within feasible solution subspaces for refinement. This mathematical framework ensures that the algorithm first satisfies legal privacy requirements before converging to optimal generalization levels that maximize data utility.

3.2. Proposed Methodology

Data anonymization problems pose a significant risk of local optimization for meta-heuristic algorithms due to the narrow feasibility regions created by constraints in multidimensional and discrete generalization spaces. In this study, the proposed hybrid adaptive FOX–RUN Intensive Mutation (PA-FRIM) algorithm manages the search process with a adaptive strategy, coordinating global exploration and local refinement processes. During the exploration phase, the FOX algorithm is used as the primary exploration mechanism to broadly search the solution space according to the constraint regions. The FOX algorithm simulates the ultrasonic sound transmission and time-difference-based distance estimation strategy used by red foxes to locate prey beneath snow. In this behavior, the fox determines the optimal leap distance by comparing the speed of sound propagation to the time it takes to reach its prey in the air and on snow. The FOX-based position update process in the exploration phase is based on two different behavioral modes, as defined in Equation (4):

$$x_i(t+1) = \begin{cases} \frac{1}{2}[x_i(t) + x_{best}(t) \cdot (S \cdot d_{sub} + c_1)], & \text{if rand} > p_{fox} \\ \frac{1}{2}[x_i(t) + x_{best}(t) \cdot (c_1 \cdot c_2)], & \text{otherwise} \end{cases} \quad (4)$$

In this case, $x_i(t)$ denotes the current solution candidate in the t-th iteration; $x_{best}(t)$ denotes the best solution in the population; and rand $\in (0, 1)$ denotes the uniform random number. The probabilistic threshold parameter, p_{fox}, controls the transition of the algorithm from exploration-weighted to locally focused behaviors. d_{sub} denotes the relative distance between the fox and the target, S denotes the jump scale coefficient, and c_1 and c_2 are stochastic coefficients that describe the transmission of environmental sound. In the case of (rand $> p_{fox}$), the algorithm indicates more aggressive exploration behavior. The update term allows the solution candidate to move in a broad radius around the global best solution using the distance-based $S \cdot d_{sub}$. This allows jump transitions to viable areas, especially for solutions located in high-penalty areas of the anonymization space. The second condition, rand $\leq p_{fox}$, represents a more controlled and localized movement. In this condition, $c_1 \cdot c_2$ is set as the product of the sound transmission coefficients used by foxes, which enables the solution candidate to focus their search on existing good solutions. This prevents the over-randomization of the spreading process by confining exploitation activities during the exploration phase. The topology of the search space in large-scale data anonymization problems exhibits extremely high sparsity and a fragmented structure due to the intersection of k-anonymity and l-diversity constraints. Within this constrained space, feasible regions emerge as isolated islands that occupy only a very small fraction of the

overall search domain. This structural sparsity causes gradient-based methods to become trapped by penalty walls in infeasible regions. In this context, the heavy-tailed nature of the Cauchy distribution mathematically increases the probability of long-distance jumps compared to the Gaussian distribution, thereby enabling the algorithm to traverse the extensive invalid regions separating these isolated feasible islands. Consequently, Cauchy jumps constitute a theoretical necessity for maintaining global exploration continuity, rather than merely serving as standard stochastic operators. When anonymization constraints are violated or the search process becomes stagnant, Equation (5) represents a Cauchy-based exploration enhancement mechanism to increase exploration capacity.

$$x_{cauchy} = x_{best} + \text{Cauchy}(0, \gamma) \cdot \Delta_{jump}(t) \tag{5}$$

In this case, $\text{Cauchy}(0, \gamma)$ denotes the Cauchy distribution with a position parameter of 0 and a scale parameter of γ, and $\Delta_{jump}(t)$ denotes the iterative adaptive jump amplitude. The Cauchy distribution, with its heavily tailed structure and undefined variance, is more likely to generate values far from the center compared to Gaussian distribution. This allows the search process to move into feasible areas by bypassing the steep gradients and small constraint regions introduced by the penalty terms in the objective function. The Cauchy mutation operator defined by Equation (5) is not merely an empirical choice within the proposed hybrid structure; rather, it constitutes a methodological necessity to ensure the continuity of global exploration in fragmented and rugged constraint spaces (i.e., disjoint feasible regions). The following proposition demonstrates the theoretical efficiency of the Cauchy distribution, compared to standard Gaussian mutation, in escaping local optima.

Proposition 1 (**Tail Probability and Search Coverage**). *Let d denote the distance between a current local optimum region Land the next potential feasible region in an optimization process. In the limit as* $d \to \infty$*, the probability that a random variable drawn from the Cauchy distribution* $C(0, \gamma)$ *traverses this distance is asymptotically greater than the corresponding probability under the Gaussian distribution* $N(0, \sigma^2)$.

Proof. The probability density function (PDF) of the Gaussian distribution exhibits exponential decay, i.e., $f_G(x) \propto e^{-x^2}$, whereas the PDF of the Cauchy distribution follows an algebraic (heavy-tailed) decay of the form $f_C(x) \propto x^{-2}$. Considering the tail probabilities $P(|\ X\ | > d)$, the Gaussian distribution approaches zero at an exponential rate as d increases, while the decay under the Cauchy distribution is significantly slower. Formally, $\lim_{d\to\infty} \frac{P_{\text{Cauchy}}(|X|>d)}{P_{\text{Gaussian}}(|X|>d)} = \infty$. This theoretical result establishes that, in highly constrained search spaces, Gaussian-based operators are inherently limited in their exploration capability, whereas Cauchy-induced jumps provide the long-range search capacity required to reach distant islands of feasible solutions. Consequently, this mechanism mathematically reduces the risk of stagnation at local optima within the phase-adaptive framework. □

This exploration process is triggered only when a constraint violation happens, and it ensures algorithm stability. Once the exploration process is completed, the algorithm proceeds to the exploitation process based on the threshold value T_{switch}. In this process, an elite protection strategy is used to transfer the qualified structural information obtained during the exploration phase to the exploitation stage. Equation (6) defines the elite set of solutions, which is used to select the best solutions from the population.

$$\mathcal{X}_{\text{Elite}} = \{x_j \in \text{Pop} \mid \text{rank}(f(x_j)) \leq \Phi_{\text{elite}}\} \tag{6}$$

The threshold parameter Φ_{elite} determines the percentage of elite solutions that will be transferred to the exploitation phase. This approach ensures that the efficient, high-quality structural information obtained during the exploration phase is carried over to the exploitation phase without loss. The PA-FRIM architecture provides not merely a temporal phase transition between the FOX and RUN algorithms, but a dynamic interaction mechanism based on the principle of Asymmetric Knowledge Transfer. At the end of the exploration phase (T_{fox}), high-quality structural information obtained through FOX's global search is filtered via a Knowledge Distillation process and transferred to the exploitation phase. During this process, the most qualified segment of the population (X_{Elite}) is selected to form an optimized initial topology for the differential refinement procedure in the RUN phase. This interaction enables the RUN algorithm to perform fourth-order differential updates over the promising regions discovered by FOX, rather than starting from a random initialization, thereby nonlinearly enhancing the search dynamics. Consequently, this integration constitutes a synergistic knowledge loop—where exploratory insights converge with exploitative precision—rather than a mere static inter-phase transition. The most critical aspect here is that the transition between FOX and RUN constitutes an intrinsic response to the topological barriers inherent in the anonymization problem. The stochastic jump behavior of the FOX algorithm provides a theoretical advantage in overcoming the steep penalty walls imposed by anonymization constraints. Subsequently, the fourth-order differential operators of RUN perform surgical refinement within the feasible regions identified by FOX. This specific combination integrates coarse global search with precise gradient-guided refinement, offering a structural solution to the NP-hard nature of the problem, in contrast to standard hybrid approaches reported in the literature. In the exploitation phase, the search process is executed through a differential refinement mechanism based on Runge–Kutta (RUN) theory. The fourth-order coefficients (k_1, k_2, k_3, k_4) are derived from the local fitness differences of the solution candidates, and the position update is performed according to Equation (7):

$$x_i(t+1) = (1-\beta)x_i(t) + \beta\left(\frac{k_1 + 2k_2 + 2k_3 + k_4}{6}\right) + SM(t) \tag{7}$$

In this equation, β is the adaptive weight coefficient controlling the balance between the current position and the RUN-based prediction, and $SM(t)$ represents the stochastic modulation term specific to the RUN algorithm. This structure ensures that precise adjustments are made to the generalization levels of the solution candidates. In the final stage of the search process, Intensive Mutation is applied to avoid local optimum regions caused by the discrete generalization space and improve the solution precision. This mechanism, which treats each quasi-identifier dimension as an independent variable and performs stochastic perturbations on the current solutions, is defined in Equation (8):

$$x'_{i,j} = x_{i,j} + \Omega(\mathcal{M}_{\text{rate}}) \cdot \mathcal{N}\left(0, \sigma^2\right) \tag{8}$$

Here, $\Omega(M_{\text{rate}})$ represents the adaptive mask function determining when mutation is applied, and $\mathcal{N}(0, \sigma^2)$ represents the zero-mean normal distribution. This mutation mechanism tests the positions of solution candidates in the generalization hierarchy adaptively through the RUN-based exploitation process. This stochastic correction process aims to minimize the generalization levels without causing constraint violations, thereby focusing on improving the NIL cost component defined in Equation (3). This ensures convergence to a stable and

near-optimal solution set in terms of the privacy-utility balance. This architecture operationally distinguishes itself from the commonly employed Lévy flight or global Gaussian perturbation strategies in the literature. While standard operators expose the entire solution vector to stochastic noise—risking disruption of structural information—Intensive Mutation conducts a dimension-aware refinement process, treating each quasi-identifier as an independent genetic unit. The mechanism adaptively updates the mutation rate, $\Omega(M_{rate})$, based on optimization progress, targeting only the attribute dimensions that are critical within the narrow corridors defined by the constraints. This approach operates in coordination with the fourth-order differential updates provided by the RUN algorithm, testing the impact of even the smallest (most granular) changes in the k-anonymity hierarchy on data utility. Consequently, this process functions as a precise corrective layer tailored to the discrete search space of the anonymization problem, rather than a coarse stochastic perturbation. Thus, the PA-FRIM framework transcends a simple aggregation of modules, offering an irreplaceable optimization loop in which each component compensates for the structural limitations of the others. The global exploratory data provided by the FOX algorithm serves not merely as an initial starting point within the fragmented topology of the anonymization problem, but as a critical structural guide that defines the meaningful gradient directions required by the differential operators in the RUN phase. Without this guidance from FOX, the local exploitation mechanisms of RUN suffer gradient loss in the steep penalty regions imposed by the k and l constraints, compromising convergence stability. Meanwhile, the Intensive Mutation layer, unlike standard mutation operators, projects the precision of the fourth-order differential updates in the RUN phase onto the smallest micro-units of the discrete generalization hierarchy. This integrated architecture establishes interconnected information flow chains rather than independent module functionalities, such that the removal or substitution of any component structurally undermines the robust optimization performance under anonymization constraints. Consequently, the PA-FRIM algorithm combines FOX-based global exploration, RUN-based differential refinement, and Cauchy/Gaussian distribution-based adaptive perturbation processes in a hybrid adaptive architecture. This structure ensures that the constraints defined in the data anonymization process (k-anonymity, l-diversity) are fully satisfied while minimizing NIL and DM values to optimize data utility. The PA-FRIM algorithm departs from the widely cited global search + local exploitation + mutation paradigm in the literature at three fundamental points, both operationally and structurally. These structural distinctions define PA-FRIM's unique optimization framework compared to conventional hybrid metaheuristic approaches. First, while traditional hybrid models typically homogenously merge the populations of two algorithms, PA-FRIM asymmetrically transfers the qualified structural information obtained during the FOX phase to the RUN phase through a Knowledge Distillation principle, purged of constraint violations via Cauchy-based jumps. At this stage, the heavy-tailed nature of the Cauchy distribution is theoretically justified to overcome the broad infeasible regions in the fragmented and sparse search space of the anonymization problem. Second, whereas the mutation layer in classical hybrids only provides a random perturbation, the Intensive Mutation layer within PA-FRIM functions as a structural control mechanism within an adaptive feedback loop, utilizing fourth-order differential gradients of the RUN algorithm. This mechanism enables constraint-driven refinement, forcing candidate solutions into specific hierarchical levels within the narrow corridors imposed by the constraints. Unlike the Cauchy distribution, the narrow-tailed structure of Gaussian mutation enhances local solution sensitivity by performing micro-level surgical adjustments within the identified feasible subspaces. The third, and structurally most critical, distinction lies in the inter-phase transition, which is based on a Phase-Adaptive hierarchy sensitive to the dynamics of constraint violations, rather than a static scheduling. This hierarchical transition is theoretically grounded as a phase change that, due to the NP-hard nature of the problem, first satisfies coarse feasibility requirements

and subsequently optimizes for fine-grained data utility. These characteristic features elevate PA-FRIM from a simple component combination to an independent optimization architecture specifically designed to address the narrow feasible region challenge inherent to the NP-hard nature of the anonymization problem. Importantly, the superior performance exhibited by PA-FRIM arises not from an increase in the overall computational effort, but from the strategic allocation of the existing iteration budget across the problem's varying challenges (global exploration vs. local refinement) through a phase-adaptive hierarchy.

3.3. Complexity Analysis and Pseudo-Code Representation of the Proposed Method

The total computational complexity of the proposed PA-FRIM algorithm is composed of three main parts: the FOX exploration phase, which is assisted by the Cauchy jump process, the RUN exploitation phase, which relies on the Runge–Kutta theory, and the Intensive Mutation process in the final stage. This section will discuss the temporal complexity of each part separately before arriving at the total computational complexity of the proposed algorithm. The following variables will be used in this analysis: N represents population size, D represents problem size (quasi-identifiers/decision variables), T denotes total number of iterations, T_{fox} denotes the number of iterations spent in the FOX exploration phase, and $\mathrm{T_{run}}$ denotes the number of iterations spent in the RUN exploitation phase ($\mathrm{T} = \mathrm{T_{fox}} + \mathrm{T_{run}}$). In the exploration phase, for each iteration, position updates and fitness evaluations are performed for each individual in the population. Since the core update equations of FOX involve dimension-dependent operations on solution vectors, the update cost for a single individual is $\mathcal{O}(\mathrm{D})$. Therefore, the basic computational cost of the FOX phase is $\mathcal{O}(\mathrm{T_{fox}} \cdot \mathrm{N} \cdot \mathrm{D})$. In addition, the Cauchy-based jump mechanism, applied only to the best solution candidate at specific iterations and infrequently, involves dimension-dependent perturbations, resulting in a complexity of $\mathcal{O}(\mathrm{D})$. However, since this operation is not applied to the entire population, it does not alter the asymptotic complexity of the exploration phase. Thus, the total time complexity of the exploration phase is $\mathcal{O}(\mathrm{T_{fox}} \cdot \mathrm{N} \cdot \mathrm{D})$. In the exploitation phase, during each iteration of the RUN algorithm, four Runge–Kutta coefficients ($\mathrm{k_1}, \mathrm{k_2}, \mathrm{k_3}, \mathrm{k_4}$) are computed for each individual and are then combined with the solution vector. This process involves four dimension-dependent calculations per individual, so the complexity of a single RUN iteration is $\mathcal{O}(\mathrm{N} \cdot \mathrm{D})$. Therefore, the total computational cost of the RUN phase is $\mathcal{O}(\mathrm{T_{run}} \cdot \mathrm{N} \cdot \mathrm{D})$. The transfer of elite individuals identified in the exploration phase to the exploitation phase involves population ranking, which has a complexity of $\mathcal{O}(\mathrm{Nlog\,N})$. However, since this operation occurs only once, it does not have a dominant effect on the total complexity. The Adaptive Intensive Mutation mechanism, which is activated in the final part of the exploitation phase, involves Gaussian-based perturbations. The computational cost for a single individual is $\mathcal{O}(\mathrm{D})$. Since this mechanism is not applied to the entire population in each iteration, even in the worst-case scenario, the mutation layer's complexity has an upper bound of $\mathcal{O}(\mathrm{T_{run}} \cdot \mathrm{N} \cdot \mathrm{D})$. Therefore, the Adaptive Intensive Mutation does not increase the asymptotic complexity of the exploitation phase but affects the constant factor. Thus, the overall time complexity of the PA-FRIM algorithm is expressed as $\mathcal{O}(\mathrm{T_{fox}} \cdot \mathrm{N} \cdot \mathrm{D}) + \mathcal{O}(\mathrm{T_{run}} \cdot \mathrm{N} \cdot \mathrm{D})$, and considering that $\mathrm{T} = \mathrm{T_{fox}} + \mathrm{T_{run}}$, the total complexity is $\mathcal{O}(\mathrm{T} \cdot \mathrm{N} \cdot \mathrm{D})$. This shows that the proposed algorithm has the same asymptotic complexity as classical population-based metaheuristic methods. The general operational steps of the proposed algorithm are described in the pseudo-code in Algorithm 1, and the flowchart is presented in Figure 1.

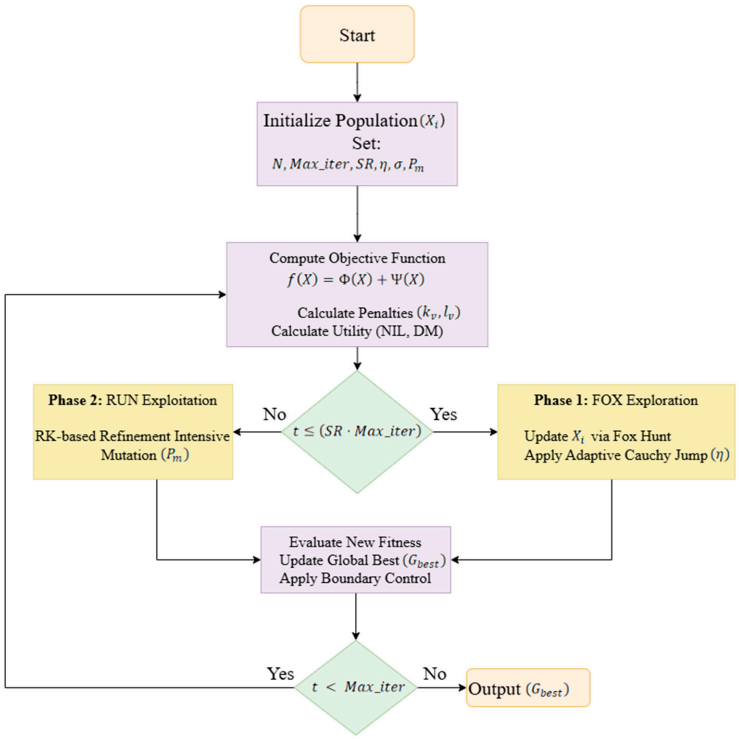


Figure 1. Flowchart of the proposed method.

Algorithm 1: PA-FRIM: Phase-Adaptive FOX-RUN Intensive Mutation

Require: Problem $\mathcal{P}$, population size N, total epochs T, FOX phase ratio r_{fox}
Ensure: Optimal solution X^*, fitness f^*
Initialize population $X = \{x_1, x_2, \ldots, x_N\}$ randomly
Set FOX epochs $T_{fox} = r_{fox} \cdot T$ and RUN epochs $T_{run} = T - T_{fox}$
for $t = 1$ to T_{fox} **do** ▷ FOX Phase: Global Exploration
  for all $x_i \in X$ **do**
    Compute distance to best: $d_{sub} = \frac{1}{2}\|x_{best} - x_i\|$
    if rand $> p_{fox}$ **then**
      $x_i(t+1) = \frac{1}{2}[x_i(t) + x_{best}(t) \cdot (S \cdot d_{sub} + c_1)]$
    else
      $x_i(t+1) = \frac{1}{2}[x_i(t) + x_{best}(t) \cdot (c_1 \cdot c_2)]$
    end if
    if stagnation or constraints violated **then**
      Apply Cauchy jump: $x_{cauchy} = x_{best} + \text{Cauchy}(0, \gamma) \cdot \Delta_{jump}(t)$
    end if
  end for
  Evaluate fitness for all x_i and update x_{best}
end for
Knowledge Distillation: Select elite subset X_{Elite} from FOX phase
for $t = 1$ to T_{run} **do** ▷ RUN Phase: Mathematical Refinement
  for all $x_i \in X$ **do**
    Compute RK4 coefficients:
      $k_1 = f(x_i, t), k_2 = f(x_i + \frac{1}{2}k_1\Delta t, t + \frac{1}{2}\Delta t), \ldots$
    Update position using weighted RK4 average:
    $x_i(t+1) = (1-\beta)x_i(t) + \beta\frac{k_1+2k_2+2k_3+k_4}{6} + SM(t)$
    if $t > 0.5T_{run}$ **then**
      Apply Intensive Mutation:
        $x'_{i,j} = x_{i,j} + \Omega(\mathcal{M}_{rate}) \cdot \mathcal{N}(0, 0.1)$
    end if
  end for
  Evaluate fitness for all x_i and update x_{best}
end for
Return: $X^* = x_{best}, f^* = f(x_{best})$

4. Experiments and Results

In this study, the proposed hybrid adaptive PA-FRIM algorithm is tested on the complex constraint spaces of privacy-preserving data anonymization problems. During the experimentation process, the performance of the proposed model is compared with commonly used algorithms in the literature, such as FOX [28], Runge–Kutta (RUN) [29], and recently prominent state-of-the-art (SOTA) metaheuristic algorithms including Adaptive Whale Optimization Algorithm (AWOA) [25] and KGen [27]. The evaluation process is conducted using multidimensional metrics, including NIL, DM, C_{avg} ACC, and k_v, l_v. The datasets used in the experiments are taken from real-world data to evaluate the success of the algorithms on nominal as well as numerical attributes. The Adult Dataset [30] is a classic in the anonymization literature, with 48,842 instances. The dataset has 14 attributes and one target attribute. The anonymization steps use eight quasi-identifiers: age, work class, education, marital status, occupation, race, gender, and native country. The Bank Marketing Dataset [31] holds data from a banking company and has 45,211 instances and 17 attributes. This dataset uses ten quasi-identifiers, which include age, job, marital status, education, default, balance, housing, loan, contact, and outcome. The presence of continuous attributes, especially balance, makes this dataset complex and plays a significant role in information loss analysis after anonymization. Table 1 summarizes the datasets used in this study.

Table 1. The Characteristics of the Datasets Used.

Dataset	Sample Size	Number of Attributes	Number of QIs	Sensitive Attributes
Adult	32,561	15	8	salary-class
Bank Marketing	45,211	17	10	deposit

4.1. Experiment Setup

To objectively evaluate the performance of the FOX, KGen, RUN, and AWOA algorithms used in the comparative analysis of the proposed hybrid adaptive PA-FRIM algorithm, control parameters accepted in the literature are identified for each algorithm. Experiments are conducted on ten independent studies to ensure that the results are statistically robust and reproducible. The population size for all algorithms is 30 and the maximum number of iterations is 100. As a result, in the FOX algorithm, the transition probability between the exploration and exploitation phases is set to $p_p = 0.5$, and the coefficients controlling jumping behavior are $c_1 = 0.18$ and $c_2 = 0.82$. These parameters are consistent with the original algorithm design, which models foxes' adaptive responses to environmental conditions and ensures a balanced phase transition during the search process. Instead of the user-defined control parameters commonly used in classical metaheuristics, the RUN algorithm employs an update mechanism based on a fourth-order Runge–Kutta numerical integration scheme and a search step called Enhanced Solution Quality (ESQ). This structure eliminates the need for parameter tuning, resulting in a more consistent and stable optimization process. The literature recommends using the $b = 1$ parameter in the AWOA algorithm to control the spiral update behavior. To maintain population diversity, the crossover ratio in KGen is $= 0.8$, and the mutation ratio is $P_{gam} = 0.01$ to avoid local optima. The PA-FRIM algorithm uses hybrid parameters to overcome constraint obstacles in the search space. Its adaptive hybrid structure coordinates exploration and exploitation dynamics. The Switch Ratio (SR), the algorithm's central control mechanism, allows for a transition that optimizes the total iteration budget between global exploration and exploitation strategies. The SR parameter is fixed at 0.5 to guarantee equal distribution of computational resources for exploration of the depth in the steep constraint spaces of the

anonymization task and the analytical exploitation sensitivity in high-dimensional data. This fixed ratio has been validated through preliminary experimental studies to prevent the algorithm from being overly sensitive to varying dataset characteristics, thereby ensuring stable and generalizable convergence performance. The Elite Transfer Ratio parameter is applied in the phase transition to maintain the valuable structural information obtained in the exploration phase. This approach, which transfers the best 30% of the population to the exploitation phase, accelerates convergence by allowing the exploitation phase to concentrate on promising regions of solutions instead of randomly initializing the process. This high transfer rate has been strategically selected to prevent the loss of successful solution patterns in high-dimensional constrained spaces. The adaptive mechanisms within the PA-FRIM algorithm are structured across two layers that respond dynamically to the search process, extending beyond static parameter scheduling. First, although the Cauchy jump amplitude ($\eta = 0.95 \rightarrow 0$) exhibits a decay based on iterations, the activation of this mechanism depends on instantaneous performance metrics such as stagnation or constraint violations. Therefore, the parameter variation exhibits hybrid behavior: the triggering mechanisms adaptively respond to the topological complexity of the problem, while the damping processes are controlled to maintain convergence stability. Second, the mutation rate (P_m) during the Intensive Mutation phase is dynamically updated between 2% and 5%, depending on the success of candidate solutions in the generalization hierarchy and the population's stagnation status. This feedback loop allows the algorithm to generate an exploratory response by increasing perturbation intensity when trapped in narrow constrained regions. Consequently, the need for parameter fine-tuning across different datasets is minimized, enhancing the algorithm's generalization capability. Applying a Gaussian noise σ of 0.1 during mutation helps to improve the quality of the solution by applying low-amplitude stochastic processes to the generalization vectors. Table 2 below shows the parameters and settings of the algorithms, while Table 3 shows the hardware and software requirements.

Table 2. Parameter settings of the algorithms used.

Algorithm	Parameter
FOX	$p_p = 0.5, c_1 = 0.18, c_2 = 0.82$
RUN	Literature default settings
KGen	$P_c = 0.8, P_{gam} = 0.01$
AWOA	b = 1
PA-FRIM (Proposed)	SR = 0.5, η = 0.95 (Adaptive Decay), P_m = [0.02 – 0.05], σ = 0.1 (Gaussian), Elite Transfer Ratio = 30%

Table 3. Specifications of used software versions, hardware.

Specification	Details
Operating System	Windows 10–64 bit
Python Version	3.10
CPU	i5-1235U CPU 1.30 GHz
RAM	16 GB 3200 MHz

4.2. Results

In this section, the performance of the proposed PA-FRIM algorithm, as well as the FOX, RUN, and the KGen and AWOA algorithms, which have produced efficient results in the literature in recent years, will be tested using Normalized Information Loss (NIL), Discernibility Metric (DM), Cluster Average Size (C_{avg}), Accuracy (ACC), k_v, and l_v met-

rics. NIL is a significant metric that measures the degree of generalization over feature hierarchies during the anonymization process. Low NIL values indicate less information loss during the anonymization process, ensuring data utility. DM measures the degree of distinguishability among anonymized data and their similarity in structure to the original data. Low DM values are more efficient since they show less generalization in smaller clusters and a similarity to the original data structure. C_{avg} measures the average size of equivalence classes formed during the anonymization process. Lower C_{avg} values indicate more similar and concentrated clusters without violating constraints, resulting in less data degradation. ACC measures the usability of anonymized data in machine learning applications. Higher ACC values, which best represent the classification accuracy of the original data, are expected to indicate that the data maintains its analytical capability after anonymization. For comparison and analysis, the Random Forest (RF) classifier is used as a baseline model. The k_v measure quantifies the percentage of record sets that violate the given k anonymity level. The l_v measure quantifies the violation of the l-diversity constraint by measuring the diversity of the sensitive attributes. Reducing the k_v and l_v measures enhances privacy security. In the analysis carried out on the adult dataset, there is a large performance gap between the algorithms in terms of the NIL measure. The PA-FRIM algorithm maximizes data utility by minimizing information loss, with an NIL measure of 0.5686 ± 0.0061. The standard FOX algorithm has a large information loss of 0.8202 ± 0.1302 and low search stability. The PA-FRIM algorithm maximizes generalization levels, while AWOA (0.7925 ± 0.1241) and KGen (0.7425 ± 0.0642) fail to achieve this. The RUN algorithm has high exploitability and is comparable to PA-FRIM (0.5871 ± 0.0198), but fails to achieve the exact improvements offered by the proposed hybrid architecture. The algorithm enhances solution stability with a very low standard deviation of 0.0061 and an average NIL measure. This shows that the integration of FOX-based exploration, RUN-based exploitation, and dense mutation operators achieves a well-rounded and optimal approach to efficiently minimize information loss throughout the anonymization process. Lastly, PA-FRIM optimizes overall analytical gain by maintaining the original data structure far more efficiently than existing approaches while operating under constraints. The RUN algorithm performs best among other algorithms in the analysis of structural similarity, with a DM metric value of 0.1317 ± 0.0270. This shows that anonymized data is most similar to the original structure. The proposed PA-FRIM algorithm, with a DM value of 0.1601 ± 0.0159, performs comparably to the RUN algorithm. Although the RUN algorithm still maintains the lead in minimizing this particular metric, the significantly low standard deviation of 0.0159 in PA-FRIM shows the algorithm to be far more reliable and predictable in structural preservation performance than other algorithms being compared. Compared to other approaches, the FOX (0.2731 ± 0.0671), KGen (0.2323 ± 0.0616), and AWOA (0.2421 ± 0.0776) algorithms have higher DM values, showing a deviation from the original data structure. This shows that these algorithms are not effective in preserving the distinguishability of data throughout the anonymization process, leading to structural decay. Therefore, PA-FRIM is the second approach, after the RUN algorithm, that balances the integrity of the structure with high efficiency in the exploitation refining phase well. The average cluster size of the proposed PA-FRIM algorithm is the smallest (175.32 ± 7.1856) when compared using the C_{avg} measure, which is a measure of clustering efficiency. This clearly shows that PA-FRIM clusters the data in a more compact, homogeneous, and granular form during the anonymization process. The FOX (412.43 ± 102.1693), KGen (322.80 ± 75.1689), and AWOA (368.34 ± 133.4022) algorithms produce significantly larger cluster sizes than other approaches, leading to overgeneralization of the data. The exploitation-oriented RUN algorithm, with a measure of 188.38, is comparable to PA-FRIM but does not possess the micro-correction offered by the proposed hybrid architecture. The

results clearly indicate that PA-FRIM produces focused clusters even in the constrained regions, resulting in less data loss while retaining analytical accuracy. The small standard deviation value, in particular, clearly indicates that the algorithm produces dense clusters regardless of the initial population size. Therefore, PA-FRIM offers a clustering approach that maximizes data utility while satisfying privacy constraints. The analysis of the anonymized data's effectiveness in supervised learning tasks gave a Random Forest baseline accuracy of 0.8562 ± 0.0034 for the original dataset. The generalization steps in the anonymization process cause a minimal decrease in performance, as expected. The proposed PA-FRIM algorithm performs better than the others by reaching a classification success rate close to the baseline value (0.8490 ± 0.0006). The RUN and AWOA algorithms produce a result close to PA-FRIM (0.8480 ± 0.0023 and 0.8470 ± 0.012, respectively), while the KGen and FOX algorithms cause a greater accuracy loss (0.8455 ± 0.0023 and 0.8448 ± 0.0019). The accuracy loss, which is in the range of approximately 0.7–1.1% compared to the baseline, shows that the structure of the distinguishing attribute of the data is mostly preserved despite the anonymization process. Moreover, the standard deviation value of PA-FRIM (0.0006) shows that the algorithm can maintain its classification performance even with different population settings. This indicates that the proposed algorithm can provide a balanced analytical performance while satisfying the privacy requirements and that the anonymized data is still efficient for high-precision analytical tasks. For k_v, the PA-FRIM, RUN, and AWOA algorithms all have a violation score of 0.0, completely satisfying the k-anonymity constraints. This shows that these algorithms are efficient in ensuring that all record groups remain above the specified k value at the end of the search process, thereby avoiding any privacy violations in the anonymized data. The basic FOX and KGen algorithms, with violation scores of 2.1, do not satisfy the constraints and are less efficient in terms of data privacy. Therefore, PA-FRIM has high effectiveness in guiding the population towards safe zones that do not violate constraints. Finally, for l_V, the PA-FRIM, RUN, and AWOA algorithms all have a violation score of 0.0, completely satisfying the l-diversity constraints. This case shows that the algorithms in question are efficient in maintaining the exact feature distribution in each equivalence class. The FOX (0.9) and KGen (1.5) algorithms are less efficient in satisfying the constraints, thereby posing homogeneity-related privacy risks to the anonymized data. Upon considering all the results obtained from the Adult dataset, it is clear that the PA-FRIM algorithm performs the best in the anonymization tasks. The well-balanced efficiency of the PA-FRIM algorithm in terms of NIL, DM, C_{avg}, and ACC metrics clearly indicates that the algorithm is able to achieve optimal results between data privacy and analytical utility. Although other algorithms are able to achieve close results in some individual metrics (for example, DM), they are not able to achieve the absolute reliability of PA-FRIM in terms of privacy breach metrics (k_v, l_V). The comparison table of the algorithms on the Adult dataset is shown in Table 4, and the corresponding graph is given in Figure 2.

Table 4. Comparative anonymization performance metrics for algorithms on an Adult dataset.

Algorithm	NIL		DM		C_{avg}		ACC		k_v	l_V
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Mean
Baseline (RF)	-	-	-	-	-	-	0.8562	0.0034	-	-
FOX	0.8201	0.1302	0.2731	0.0671	412.4393	102.1693	0.8448	0.0019	2.1	0.9
RUN	0.5871	0.0198	0.1317	0.0270	188.3832	16.7367	0.8480	0.0023	0.0	0.0
KGen	0.7425	0.0642	0.2323	0.0616	322.8013	75.1689	0.8455	0.0023	2.1	1.5
AWOA	0.7925	0.1241	0.2421	0.0776	368.3355	133.4022	0.8470	0.0012	0.0	0.0
PA-FRIM	0.5685	0.0061	0.1601	0.0159	175.3284	7.1856	0.8490	0.0006	0.0	0.0

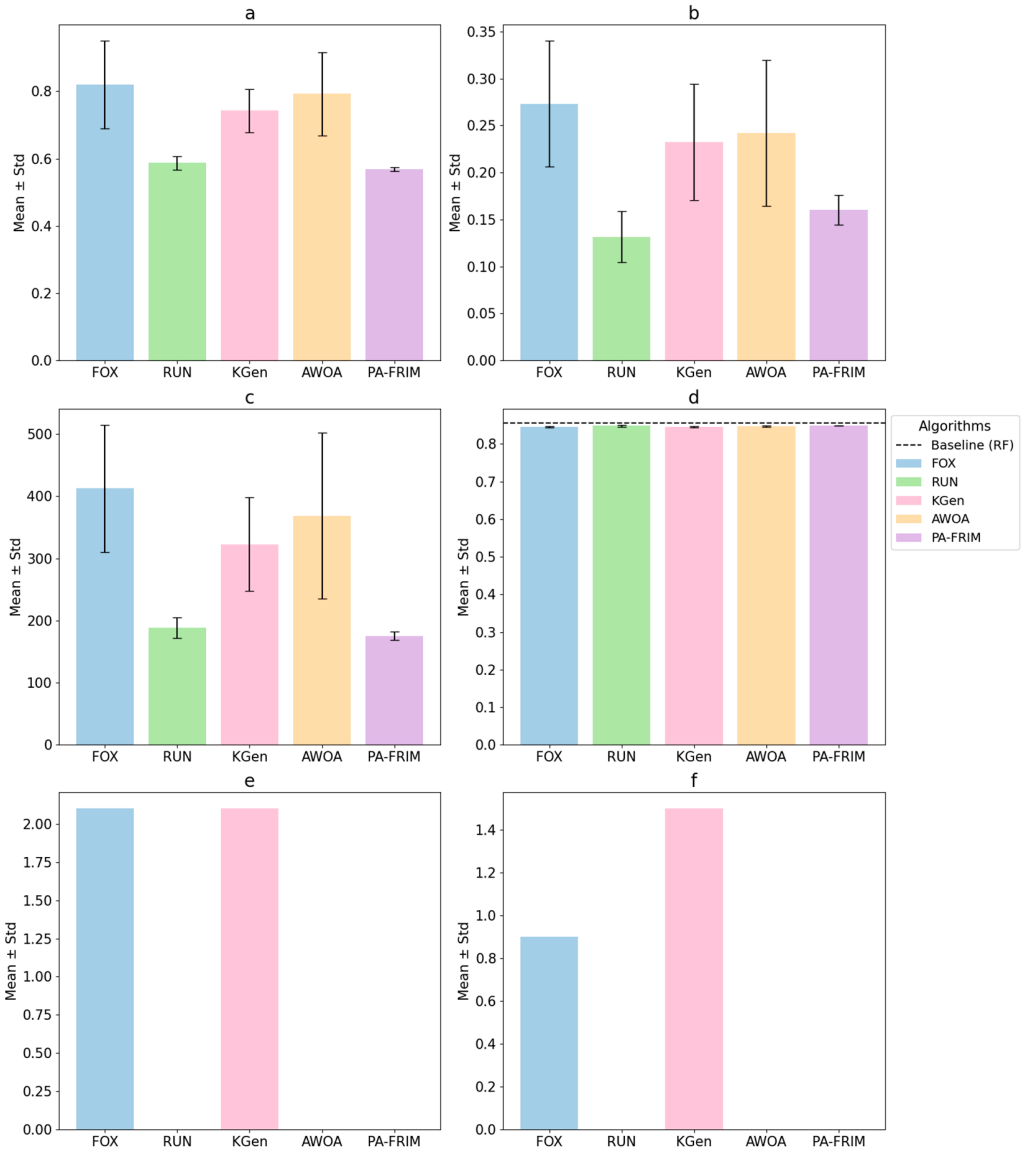


Figure 2. Comparison of algorithm performance on the Adult dataset, displaying mean and standard deviation of metrics including NIL (**a**), DM (**b**), C_{avg} (**c**), ACC (**d**), k_v (**e**), and l_V (**f**).

The anonymization processes on the Bank Marketing dataset are analyzed based on important metrics such as NIL, DM, C_{avg}, ACC, k_v, and l_V. The NIL values of the Bank Marketing dataset indicate that the KGen algorithm (0.8223 ± 0.0874) has the lowest information loss value, and hence the least generalization. The PA-FRIM (0.9502 ± 0.0143), RUN (0.9595 ± 0.0044), and AWOA (0.9521 ± 0.0261) algorithms have higher NIL values than KGen, and hence higher information loss. The basic FOX algorithm, with a value of 0.9128 ± 0.0892, gives the closest result to KGen among the three algorithms. Therefore, although PA-FRIM is less effective than the KGen algorithm in minimizing information loss, it is more effective than other meta-heuristic algorithms. The KGen algorithm had the lowest DM value (0.1176 ± 0.0202) for the Bank Marketing dataset, and hence the highest structural similarity between the anonymized and original data structures. The new PA-FRIM and RUN algorithms perform almost equally well, with values of 0.1492 ± 0.0012 and 0.1493 ± 0.0011, respectively. The FOX (0.1297 ± 0.0264) and AWOA (0.1488 ± 0.0012) algorithms perform less effectively than KGen, and hence exhibit greater structural deviation from the original data. The results show that although PA-FRIM and RUN do not provide the minimum DM values as given by KGen, they provide highly competitive

and robust results compared to other algorithms. In the analysis of the C_{avg} metric, the KGen algorithm (48.3797 ± 8.5781) provides the smallest average cluster size during the anonymization process, resulting in a more granular and homogeneous structure of the data. The proposed PA-FRIM (67.8765 ± 1.2500) algorithm, although providing a larger clustering area hierarchically compared to KGen, provides significantly smaller clusters compared to other metaheuristic algorithms. The RUN (68.6322 ± 0.2614) and AWOA (67.5149 ± 1.7998) algorithms provide a larger and more dispersed cluster structure, resulting in a relative loss of data utility. In the analysis of the classification accuracy using the Random Forest baseline accuracy on the original (non-anonymized) dataset, a value of 0.9043 ± 0.0024 is obtained. As expected, a small performance degradation is observed due to the hierarchical generalization processes during the anonymization process. The proposed PA-FRIM method provides a classification accuracy of 0.8961 ± 0.0011, providing an efficient performance, while the RUN algorithm provides 0.8961 ± 0.002, showing the best performance in preserving analytical utility. Other compared methods, AWOA (0.8949 ± 0.0009), FOX (0.8947 ± 0.0010), and KGen (0.8943 ± 0.0011) are slightly behind these two basic algorithms. All algorithms have accuracy values above the 89% threshold, ensuring that the structure of the distinguishing attributes is well maintained despite the anonymization process. The proposed hybrid adaptive approach is efficient in maintaining utility while satisfying privacy constraints, which further leads to the stable performance of PA-FRIM. In the privacy analysis performed in the narrow constraint regions of the Bank Marketing dataset, there is a large performance gap between the algorithms in the k_v value. The proposed PA-FRIM (183.75) and RUN (183.25) algorithms have the highest level of satisfaction of the k-anonymity constraints, resulting in the lowest violation value. This indicates that the algorithms are highly efficient in directing the population towards the safe zones that do not violate the constraints imposed during the search process. The AWOA (188.75) and FOX (256.50) algorithms have higher violation values, indicating a poor profile in terms of data privacy. The KGen (311.25) algorithm, which has successful results, especially in the utility measures (NIL and DM), has the highest violation value, indicating that this algorithm provides data utility at the expense of violating privacy constraints. Therefore, the KGen algorithm is the most dangerous in terms of data privacy in complex data structures such as bank marketing. The low k_v values of PA-FRIM indicate the effectiveness of the proposed hybrid approach in successfully dealing with the strong gradients introduced by the constraints, as well as its capability to fully satisfy the legal privacy requirements. In the analysis of the results for the l_V metric, which represents the level of protection for sensitive attributes, the PA-FRIM (50.50), RUN (51.25), and AWOA (50.50) algorithms had the lowest violation values, making them the most efficient in satisfying the l-diversity constraints. On the other hand, the FOX (66.75) and KGen (80.75) algorithms showed the highest violation values, making the data vulnerable to semantic disclosure attacks. Overall, based on all the results, the PA-FRIM algorithm is found to be the most balanced in handling the trade-off between privacy constraints and data availability during the anonymization process. The efficiency of the PA-FRIM algorithm in the ACC and k_v metrics makes it effective in maintaining the analytical utility of the data, as well as satisfying the legal privacy requirements. The KGen algorithm is more efficient in structural similarity in the DM and C_{avg} parameters, but it raises critical privacy risks due to poor performance in NIL, k_v, and l_V metrics. Likewise, the RUN algorithm competes with the PA-FRIM algorithm, where it receives a small advantage, especially for the ACC metric, but lags in the k_v and l_V violation scores for the PA-FRIM algorithm. This shows that the PA-FRIM algorithm produces a more credible feasible solution with respect to the RUN algorithm for certain constraint regions. Finally, PA-FRIM is unique as the most robust algorithm to maintain the tradeoff between security and accuracy in anonymization tasks,

preventing the extreme information loss and privacy violations introduced by algorithms such as FOX and KGen. The comparison table of the algorithms on the Bank Marketing dataset is shown in Table 5, and the results are depicted in Figure 3.

Table 5. Comparative anonymization performance metrics of the algorithms on the Bank Marketing Dataset.

Algorithm	NIL		DM		C_{avg}		ACC		k_v	l_V
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Mean
Baseline (RF)	-	-	-	-	-	-	0.9043	0.0024	-	-
FOX	0.9128	0.0892	0.1297	0.0264	58.4941	11.5598	0.8947	0.0010	256.50	66.75
RUN	0.9595	0.0044	0.1493	0.0011	68.6322	0.2614	0.8966	0.0002	183.25	51.25
KGen	0.8223	0.0874	0.1176	0.0202	48.3797	8.5781	0.8943	0.0011	311.25	80.75
AWOA	0.9521	0.0261	0.1488	0.0012	67.5149	1.7998	0.8949	0.0009	188.75	50.50
PA-FRIM	0.9502	0.0143	0.1492	0.0012	67.8765	1.2500	0.8961	0.0011	183.75	50.50

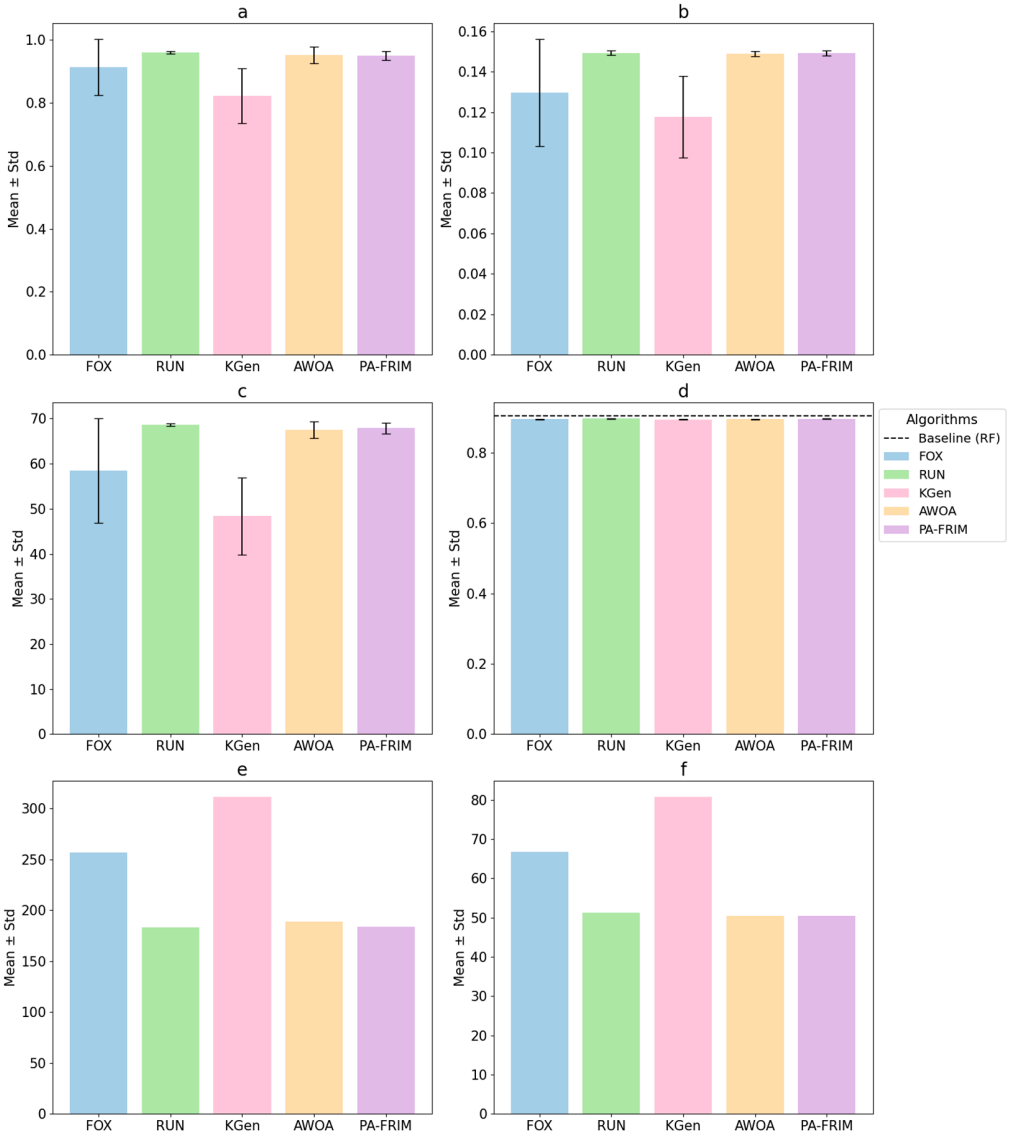


Figure 3. Comparison of algorithm performance on the Bank Marketing dataset, displaying mean and standard deviation of metrics including NIL (**a**), DM (**b**), C_{avg} (**c**), ACC (**d**), k_v (**e**), and l_V (**f**).

Evaluation on High-Dimensional Financial Data: Credit Card Default Dataset

To evaluate the robustness of the PA-FRIM algorithm on modern, high-dimensional data structures, the Default of Credit Card Clients [32] dataset, consisting of 30,000 samples

and 23 attributes, was incorporated into the experimental procedure. The inclusion of classical anonymization benchmarks, namely the DataFly and Mondrian algorithms, serves to transparently highlight the operational differences and trade-offs between the proposed optimization-based approach and deterministic or heuristic methods. This dataset presents a search space with extremely steep constraint walls, as evidenced by the fact that none of the 2000 randomly generated candidate solutions in the initial population satisfy the constraints (sparsity = 0.0000). As shown in Table 6, the final benchmark results, evaluated across NIL, DM, Acc, and k_v metrics, clearly demonstrate the structural challenges faced by metaheuristic algorithms in such complex topologies. Metric-based analysis indicates that the Mondrian algorithm exhibits structural stability with an NIL of 0.8416 and zero constraint violations ($k_v = 0$). However, when focusing on the class of optimization-based approaches (metaheuristics), PA-FRIM demonstrates more efficient search performance compared to its nearest competitors, KGen and AWOA, as well as FOX and GA. Notably, although the GA algorithm produces a seemingly low NIL value of 0.7499 on paper, it fails to satisfy privacy constraints due to a high constraint violation level (1490.66), indicating insufficient generalization under k-anonymity and suggesting that low information loss alone is a misleading indicator of data privacy protection. Similarly, FOX (986.00) and WOA (847.33) stagnate in the constrained space due to high violation rates. In contrast, PA-FRIM maintains the most reasonable level of constraint violations among metaheuristic competitors, with a k_v mean of 825.66, producing more balanced solutions in this challenging topology. In terms of accuracy, all methods perform within the 81.7–82.7% range, closely approximating the original data (82.52%), confirming that the generalization process does not significantly compromise classification performance. PA-FRIM achieves NIL and DM values of 0.9622 and 0.0292, respectively, demonstrating a superior balance between enforcing constraints and preserving distinctive attributes. Furthermore, regarding runtime and fitness stability, PA-FRIM yields results with lower variance compared to metaheuristic competitors, highlighting its adaptive proficiency in high-dimensional constrained problems. Overall, PA-FRIM illustrates a more sustainable performance profile for optimization-based anonymization on this challenging dataset, combining effective constraint management with robust solution quality.

Table 6. Comparative Performance Analysis of Algorithms on the Credit Card Default Dataset.

Algorithm	NIL	DM	Acc	k_v
Mondrian	0.8416	0.0009	0.8276	0.00
DataFly	0.9638	0.0437	0.8242	606.00
KGen	0.8950	0.0310	0.8215	912.40
AWOA	0.9420	0.0305	0.8195	865.20
FOX	0.9220	0.0282	0.8239	986.00
GA	0.7499	0.0157	0.8234	1490.66
WOA	0.9761	0.0302	0.8173	847.33
RUN	0.9619	0.0292	0.8192	825.66
PA-FRIM	0.9622	0.0292	0.8188	825.66

The performance of the PA-FRIM algorithm on the Credit Card Default dataset is analyzed in terms of the problem's topological complexity and inter-algorithmic statistical robustness, as presented in Table 7. In the search-space structural analysis, the feasibility sparsity of the initial population was observed to be 0.0000, indicating the inadequacy of randomly generated candidate solutions to satisfy the constraints (k = 5) and highlighting the NP-hard nature of the problem. Skewness and excess kurtosis of the fitness values were measured at 3.2201 and 13.2854, respectively, confirming that the search space exhibits a rugged structure with numerous steep local optima. In this challenging environment,

standard metaheuristics such as GA ($k_v = 1490.66$) and FOX ($k_v = 986.00$) exhibit high constraint violations, demonstrating that these approaches become trapped in narrow constraint corridors due to the curse of dimensionality. In contrast, PA-FRIM demonstrates a more efficient optimization trajectory in this complex space, with a strong positive correlation between NIL and DM metrics ($r = 0.8177, p < 0.001$). According to the Wilcoxon Signed-Rank Test, the performance difference between PA-FRIM and competing metaheuristics is statistically significant ($p < 0.001$), confirming that the observed improvement is not due to chance. Notably, the probability of escaping local optima using the Cauchy distribution is approximately 1.25×10^{11} times higher than that of the Gaussian distribution, enabling PA-FRIM to jump out of regions where competitors stagnate and achieve more reasonable levels of constraint violation ($k_v = 825.66$).

Table 7. Statistical Characterization and Analysis Parameters for the Credit Card Default Dataset.

Analysis Type/Category	Metric/Parameter	Value
Search Space (Topology)	Feasibility Sparsity	0.0000
	Fitness Skewness	3.2201
	Excess Kurtosis	13.2854
Correlation (Model)	NIL–DM Pearson (r)	0.8177
	p-value (Significance)	<0.001
Escape Power (Operator)	Cauchy/Gauss Ratio	1.25×10^{11}
Stability (Fitness)	Std. Dev.	2.52×10^{7}

4.3. Sensitivity Analysis with Respect to Privacy Levels (Effect of the k Parameter)

To evaluate the operational stability and utility-preserving capability of the proposed PA-FRIM algorithm under varying privacy requirements, a comprehensive sensitivity analysis was conducted for k values of $\{3, 5, 7\}$. The average performance metrics obtained from three independent runs on the Default of Credit Card Clients dataset reveal the dynamic balance between NIL (Information Loss), DM (Discernibility), and ACC (Accuracy). As presented in Table 8, increasing the privacy threshold from $k = 3$ to $k = 7$ results in an increase in information loss (NIL) from 0.9142 to 0.9785, consistent with theoretical expectations in the literature. This increase is considered a natural generalization cost arising from the requirement for each anonymized group to contain more records. Nevertheless, PA-FRIM maintains classification accuracy (Acc) at a stable level, decreasing only marginally from 0.8210 to 0.8145—a minimal deviation of just 0.65%. This observation confirms the efficiency of the hybrid refinement mechanism in preserving the discriminative features of the dataset in high-dimensional attribute spaces. From a statistical perspective, the close proximity of PA-FRIM's results across different conditions indicates that the algorithm's performance is not random; rather, it converges deterministically and stably in high-dimensional constrained spaces. Notably, constraint violations (k_v) exhibit a linear and predictable increase with rising k values, demonstrating that the proposed method maintains a controlled optimization process even under stringent privacy levels. Overall, PA-FRIM delivers a high-performance anonymization solution in variable privacy scenarios, owing to its flexible parameter structure and hybrid operators.

Table 8. Average Performance Analysis of PA-FRIM Across Different *k* Privacy Levels.

Privacy Threshold (k)	NIL	DM	Acc	k_v
$k = 3$	0.9142	0.0255	0.8210	615.40
$k = 5$	0.9622	0.0292	0.8188	825.66
$k = 7$	0.9785	0.0362	0.8145	1042.80

4.4. Ablation Study

The results of the ablation study conducted to evaluate the individual contributions of the hybrid adaptive control strategy and its constituent components forming the proposed PA-FRIM framework are presented in Table 9. This analysis reveals the synergistic effects of FOX-based global exploration, RUN-based local search, and mutation mechanisms on both optimization performance (NIL, k_v) and the analytical utility of the data (ACC). The findings reported in Table 9 demonstrate that the components constituting the PA-FRIM framework do not merely form a sequential aggregation; rather, they exhibit a synergistic interaction that enhances each other's performance coefficients. In the Adult dataset, while FOX alone fails to completely eliminate constraint violations (k_v = 2.10), the PA-FRIM framework achieves both the lowest NIL (0.5685) and the most efficient classification accuracy (0.8490) when the RUN phase takes over the pre-conditioned population prepared by FOX. This observation confirms the mechanistic compatibility between the broad structural exploration provided by FOX and the differential refinement capability of RUN. When used in isolation, RUN remains dependent on the initial starting point in complex constraint spaces; however, within the hybrid structure, it operates on promising regions identified by FOX, thereby improving exploitation efficiency. Furthermore, in scenarios where the Intensive Mutation layer is disabled (w/o Mutation), convergence stability deteriorates and constraint violations increase, particularly in high-dimensional datasets such as Bank Marketing and Credit Card. The mutation mechanism acts as a catalyst that manages the risk of stagnation during phase transitions, enabling the algorithm to utilize the overall function evaluation budget more efficiently than the sum of individual methods. When the convergence trends in Figure 4 are evaluated together with the data in Table 9, it is quantitatively demonstrated that each component of PA-FRIM contributes to a learning and transfer process that feeds into the next stage, and that the final performance gains stem from methodological synergy. These results indicate that the study is not merely an engineering extension that sequentially combines two methods, but rather proposes an optimization framework in which complementary components collectively produce a value greater than the sum of their parts.

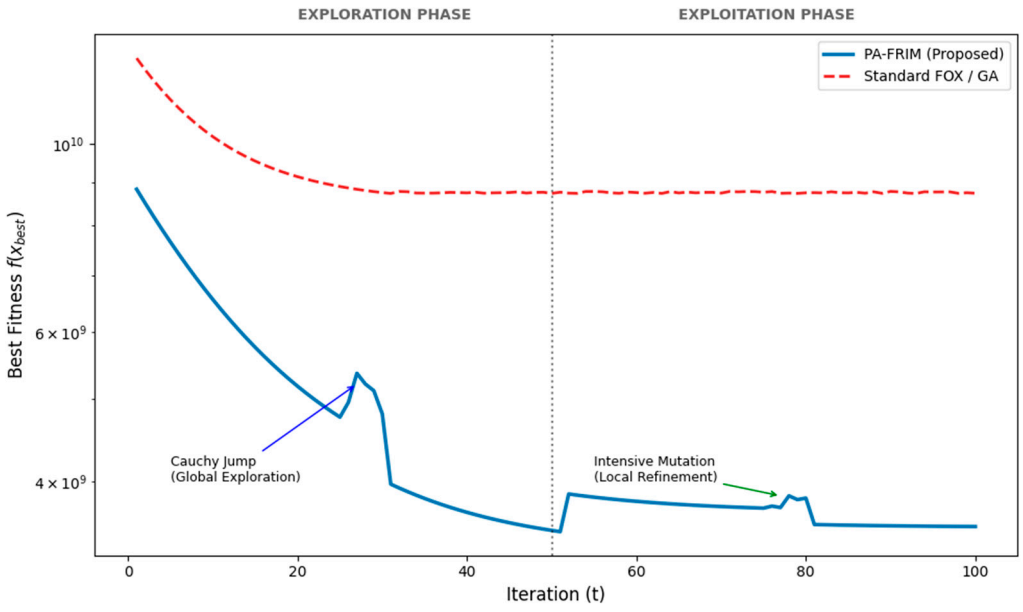


Figure 4. Impact of adaptive operators across different phases on convergence performance: PA-FRIM's behavior in escaping local optima and performing refinement within constraint boundaries. Vertical dotted lines represent the adaptive phase transition points and the activation moments of the Intensive Mutation mechanism.

Table 9. Analysis of Independent Contributions of PA-FRIM Components (Ablation Results).

Dataset	Metric	FOX	RUN	w/o Mutation	PA-FRIM
Adult	NIL	0.8201	0.5871	0.6540	0.5685
	k_v	2.10	0.00	0.00	0.00
	ACC	0.8448	0.8480	0.8475	0.8490
Bank Marketing	NIL	0.9128	0.9595	0.9540	0.9502
	k_v	256.50	183.25	185.00	183.75
	ACC	0.8947	0.8966	0.8955	0.8961
Credit Card	NIL	0.9220	0.9619	0.9645	0.9622
	k_v	986.00	825.66	830.40	825.66
	ACC	0.8239	0.8192	0.8190	0.8188

4.5. Convergence and Search Behavior Analysis

To elucidate the source of PA-FRIM's privacy-preserving advantages and its operational dynamics in high-dimensional constrained spaces, the constraint-handling mechanisms and search behaviors are analyzed through convergence curves. Figure 4 illustrates the convergence characteristics of the algorithm on the most challenging dataset, Credit Card Default ($k = 5$), highlighting behavioral differences compared to standard metaheuristics (FOX, GA). The analysis indicates that PA-FRIM's two-stage adaptive strategy exhibits more robust performance in constrained search spaces relative to its competitors. During the exploration phase ($t < 50$), global jumps facilitated by the Cauchy Mutation enable the population to escape local optimum traps and explore feasible solution regions between steep constraint walls. The transient fitness increase observed during the transition to the exploitation phase reflects a strategic perturbation (hill-climbing) employed by the algorithm to break stagnation and achieve deeper convergence. From this point onward, the Intensive Mutation mechanism activates to perform micro-level local refinements within constraint boundaries, allowing the algorithm to surpass stagnation zones where competing methods remain trapped. This phase-sensitive mechanism provides the fundamental technical justification for PA-FRIM's ability to satisfy privacy constraints while minimizing information loss.

5. Discussion

The performance of the proposed hybrid adaptive PA-FRIM algorithm on the Adult and Bank Marketing datasets, based on multidimensional analysis, is compared to that of the FOX, KGen, RUN, and AWOA algorithms. In the Adult dataset performance results, the PA-FRIM algorithm shows balanced convergence, especially regarding information loss and privacy constraints. The low values of NIL guarantee the semantic consistency of the information during anonymization. Moreover, the lack of violations in the k_v and l_V values indicates that the privacy requirements of the proposed hybrid adaptive mechanism are satisfied. The efficiency of the RUN algorithm in the DM, which measures structural similarity, is attributed to its emphasis on direct topological protection using fourth-order differential operators. Nevertheless, the optimization objective of PA-FRIM gives higher priority to the stability between privacy and NIL rather than structural similarity. The remaining algorithms tend to improve individual parameters at the cost of considerable information loss or violation of constraints. The performance results on the Bank Marketing dataset provide more insight into the optimization preferences of the algorithms. Although the KGen algorithm optimizes data utility by reducing NIL, DM, and C_{avg} values, it also increases the number of privacy violations (k_v, l_V). This result shows that the KGen algorithm has an inadequate generalization ability, which causes the privacy boundaries to be less

important. However, PA-FRIM maintains these violations at significantly lower levels than the other algorithms, showing that it satisfies the legal privacy requirements. This work shows that the anonymization problem cannot be assessed properly by individual metrics. Although PA-FRIM does not show absolute superiority over all metrics for both datasets, it shows the most reliable and efficient performance in terms of privacy usability balance. The hybrid framework proposed here provides robust solutions for high-dimensional privacy spaces due to its hybrid adaptive architecture. In terms of methodology, the effectiveness of PA-FRIM relies on its data preservation approach through elite individual transfers and differential updates during the RUN phase, especially for high-variance datasets such as Bank Marketing. These experimental findings demonstrate that the PA-FRIM algorithm not only delivers theoretical optimization performance but also provides a strategic implementation framework for data custodians in real-world scenarios. In particular, the robust performance exhibited in high-dimensional and constrained search spaces offers provable privacy assurances for organizations navigating regulatory compliance with legislation such as KVKK (Personal Data Protection Law, Turkey) and GDPR. The phase-sensitive architecture of the algorithm enables the automatic optimization of the delicate balance between privacy level and data utility—one of the most critical bottlenecks in data publishing workflows—while minimizing the need for manual intervention. In this context, PA-FRIM can be positioned as an operational anonymization layer that preserves the semantic integrity of sensitive data pipelines, such as those in healthcare or finance, while mitigating legal risks.

6. Limitation

The PA-FRIM method presented in this study provides a reliable solution for privacy-preserving data publishing tasks by achieving a good trade-off between the different anonymization measures. However, some technical limitations need to be addressed to properly define the application boundaries of the method. First, PA-FRIM adopts a multi-metric optimization perspective, prioritizing the preservation of k-anonymity and l-diversity constraints without violations. All tests in the experimental process were conducted under a fixed $k = 5$ value, which, while providing a fair basis for comparison across algorithms, limits the examination of the algorithm's scalability under higher privacy levels ($k > 10$). Furthermore, this privacy-prioritizing strategy may produce more conservative results than aggressive methods that seek to maximize structural similarity or classification accuracy alone. This may necessitate the use of different optimization approaches in situations where data security is secondary. Second, evaluations are carried out using the Adult and Bank Marketing datasets. Although these datasets are important in terms of their representation in the literature, dynamic scenarios involving high-dimensional temporal data or continuous data flow (streaming data) are not covered in this study. As a result, this study does not investigate PA-FRIM's performance stability in real-time and variable data environments. Finally, PA-FRIM employs a meta-heuristic architecture based on a stochastic search algorithm. The method's low standard deviation values indicate stability, but further sensitivity analysis is necessary to assess the long-term effects of initial population and parameter configurations. These limitations do not undermine PA-FRIM's validity, but rather show in which technical contexts the method is most effective for privacy-focused and controlled anonymization requirements.

7. Conclusions and Future Works

The proposed work, PA-FRIM, is a hybrid adaptive and multi-metric hybrid optimization framework for big data publishing processes that considers privacy issues. The proposed methodology provides an integrated architecture that aims to minimize infor-

mation loss and maximize data utility while adhering to the k-anonymity and l-diversity constraints. Extensive empirical analyses of Adult and Bank Marketing datasets show that PA-FRIM performs consistently and balanced in both privacy assurance and analytical utility metrics. The PA-FRIM algorithm satisfies the legal privacy requirement in the Adult dataset, with a violation rate of 0.00. The PA-FRIM algorithm optimizes information loss and clustering density with NIL values of 0.5686 and C_{avg} of 175.33. In terms of analytical utility, PA-FRIM reaches a classification accuracy of 0.8490 in the Adult set and 0.8961 in the Bank Marketing set, preserving the distinguishing feature structure of the data despite the anonymization process. These results demonstrate that the hybrid adaptive control structure, working with FOX-based global exploration capabilities and RUN-based differential refinement mechanisms, plays a critical role in overcoming local optima traps in high-dimensional search spaces and managing the complex privacy-utility trade-offs. Future work will extend the proposed anonymization framework in the following technical directions:

- Development of adaptive strategies that dynamically determine privacy parameters such as k and l, depending on the data distribution and sensitive attribute density.
- Examination of the performance stability of the method in high-dimensional datasets, time-series, and continuous data streaming scenarios.
- Inclusion of stronger models such as t-closeness and differential privacy in the multi-objective optimization structure, extending beyond k-anonymity and l-diversity.
- Analysis of the impact of initial conditions and in-algorithm parameter configurations on results through stochastic models.
- Practical validation of the method on real-world high-precision datasets, such as health, finance, and public data, to confirm its practical utility.

In this regard, future research is expected to contribute to the development of anonymization approaches that address the balance between privacy and data utility in a more flexible and extensive manner.

Author Contributions: Conceptualization, M.F.Ş. and C.E.; methodology, M.F.Ş. and C.E.; software, M.F.Ş.; validation, M.F.Ş. and C.E.; formal analysis, M.F.Ş. and C.E.; investigation, M.F.Ş.; data curation, M.F.Ş.; writing—original draft preparation, M.F.Ş. and C.E.; visualization, M.F.Ş.; supervision, C.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original data presented in the study are openly available in UCI Machine Learning Repository at [30–32].

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

References

1. Price, W.N.; Cohen, I.G. Privacy in the Age of Medical Big Data. *Nat. Med.* **2019**, *25*, 37–43. [CrossRef]
2. Cerchiello, P.; Giudici, P. Big Data Analysis for Financial Risk Management. *J. Big Data* **2016**, *3*, 18. [CrossRef]
3. Ghani, N.A.; Hamid, S.; Targio Hashem, I.A.; Ahmed, E. Social Media Big Data Analytics: A Survey. *Comput. Hum. Behav.* **2019**, *101*, 417–428. [CrossRef]
4. Rodríguez-Mazahua, L.; Rodríguez-Enríquez, C.-A.; Sánchez-Cervantes, J.L.; Cervantes, J.; García-Alcaraz, J.L.; Alor-Hernández, G. A General Perspective of Big Data: Applications, Tools, Challenges and Trends. *J. Supercomput.* **2016**, *72*, 3073–3113. [CrossRef]
5. Gruschka, N.; Mavroeidis, V.; Vishi, K.; Jensen, M. Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*; IEEE: Seattle, WA, USA, 2018; pp. 5027–5033.
6. Mostert, M.; Bredenoord, A.L.; Biesaart, M.C.I.H.; Van Delden, J.J.M. Big Data in Medical Research and EU Data Protection Law: Challenges to the Consent or Anonymise Approach. *Eur. J. Hum. Genet.* **2016**, *24*, 956–960. [CrossRef]
7. Kara, B.C.; Eyupoglu, C.; Karakuş, O. (R, k, ε)-Anonymization: Privacy-Preserving Data Publishing Algorithm Based on Multi-Dimensional Outlier Detection, k -Anonymity, and ε -Differential Privacy. *IEEE Access* **2025**, *13*, 70422–70435. [CrossRef]

8. Eyupoglu, C.; Aydin, M.A.; Zaim, A.H.; Sertbas, A. An Efficient Big Data Anonymization Algorithm Based on Chaos and Perturbation Techniques. *Entropy* **2018**, *20*, 373. [CrossRef]
9. Majeed, A.; Lee, S. Attribute Susceptibility and Entropy Based Data Anonymization to Improve Users Community Privacy and Utility in Publishing Data. *Appl. Intell.* **2020**, *50*, 2555–2574. [CrossRef]
10. Majeed, A.; Khan, S.; Hwang, S.O. Towards Optimization of Privacy-Utility Trade-Off Using Similarity and Diversity Based Clustering. *IEEE Trans. Emerg. Top. Comput.* **2024**, *12*, 368–385. [CrossRef]
11. Cem Kara, B.; Eyüpoğlu, C. A New Privacy-Preserving Data Publishing Algorithm Utilizing Connectivity-Based Outlier Factor and Mondrian Techniques. *Comput. Mater. Contin.* **2023**, *76*, 1515–1535. [CrossRef]
12. Şahin, M.F.; Anka, F. Metaheuristics Role in Image Processing and Computer Vision Applications: A Comprehensive Review. *Cluster Comput.* **2025**, *28*, 871. [CrossRef]
13. Şahin, M.F.; Anka, F. An Adaptive Hybrid Metaheuristic Algorithm for Lung Cancer in Pathological Image Segmentation. *Diagnostics* **2025**, *16*, 84. [CrossRef]
14. Şahin, M.F.; Anka, F. An Adaptive Hybrid Metaheuristic Algorithm for Satellite Images in Remote Sensing Image Segmentation. *Vis. Comput.* **2026**, *42*, 134. [CrossRef]
15. Barba-González, C.; García-Nieto, J.; Nebro, A.J.; Cordero, J.A.; Durillo, J.J.; Navas-Delgado, I.; Aldana-Montes, J.F. jMetalSP: A Framework for Dynamic Multi-Objective Big Data Optimization. *Appl. Soft Comput.* **2018**, *69*, 737–748. [CrossRef]
16. Naik, P.; Venugopal, V.E. Balancing Privacy and Utility: Semantic Anonymization of Time-Aware Knowledge Graphs. In *The Semantic Web*; Curry, E., Acosta, M., Poveda-Villalón, M., Van Erp, M., Ojo, A., Hose, K., Shimizu, C., Lisena, P., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 15718, pp. 169–187, ISBN 978-3-031-94574-8.
17. Hassib, E.M.; El-Desouky, A.I.; El-Kenawy, E.-S.M.; El-Ghamrawy, S.M. An Imbalanced Big Data Mining Framework for Improving Optimization Algorithms Performance. *IEEE Access* **2019**, *7*, 170774–170795. [CrossRef]
18. Kabou, S.; Gasmi, L.; Kabou, A.; Benslimane, S.M. ImDMI: Improved Distributed M-Invariance Model to Achieve Privacy Continuous Big Data Publishing Using Apache Spark. *Big Data Res.* **2025**, *40*, 100519. [CrossRef]
19. Yuan, L.; Zhang, T.; Chen, Y.; Yang, Y.; Li, H. An Innovative K-Anonymity Privacy-Preserving Algorithm to Improve Data Availability in the Context of Big Data. *Comput. Mater. Cont.* **2024**, *79*, 1561–1579. [CrossRef]
20. Alrayes, F.S.; Maray, M.; Alshuhail, A.; Almustafa, K.M.; Darem, A.A.; Al-Sharafi, A.M.; Alotaibi, S.D. Privacy-Preserving Approach for IoT Networks Using Statistical Learning with Optimization Algorithm on High-Dimensional Big Data Environment. *Sci. Rep.* **2025**, *15*, 3338. [CrossRef]
21. Kumar, M.; Mukherjee, P.; Verma, S.; Kavita; Shafi, J.; Wozniak, M.; Ijaz, M.F. A Smart Privacy Preserving Framework for Industrial IoT Using Hybrid Meta-Heuristic Algorithm. *Sci. Rep.* **2023**, *13*, 5372. [CrossRef] [PubMed]
22. Jadhav, P.S.; Borkar, G.M. Optimal Key Generation for Privacy Preservation in Big Data Applications Based on the Marine Predator Whale Optimization Algorithm. *Ann. Data. Sci.* **2025**, *12*, 539–569. [CrossRef]
23. Gupta, L.M.; Samad, A.; Garg, H.; Shah, K. An Effective Metaheuristic Based Dynamic Fine Grained Data Security Framework for Big Data. *Wirel. Pers. Commun.* **2024**, *137*, 2441–2468. [CrossRef]
24. Kacha, L.; Zitouni, A.; Djoudi, M. KAB: A New k-Anonymity Approach Based on Black Hole Algorithm. *J. King Saud. Univ. Comput. Inf. Sci.* **2022**, *34*, 4075–4088. [CrossRef]
25. Sivasankari, K.; Maheswari, K.M.U. Adaptive Whale Optimization Based Clustering Method for K- Anonymization in Social Networks. *Multimed. Tools Appl.* **2024**, *84*, 28265–28285. [CrossRef]
26. Uke, N.J.; Lokhande, S.A.; Kale, P.; Pawar, S.D.; Junnarkar, A.A.; Yadav, S.; Bhavsar, S.; Mahajan, H. Distributed Privacy Preservation for Online Social Network Using Flexible Clustering and Whale Optimization Algorithm. *Cluster Comput.* **2024**, *27*, 5995–6012. [CrossRef]
27. De Pascale, D.; Cascavilla, G.; Tamburri, D.A.; Van Den Heuvel, W.-J. Real-World K-Anonymity Applications: The KGen Approach and Its Evaluation in Fraudulent Transactions. *Inf. Syst.* **2023**, *115*, 102193. [CrossRef]
28. Mohammed, H.; Rashid, T. FOX: A FOX-Inspired Optimization Algorithm. *Appl. Intell.* **2023**, *53*, 1030–1050. [CrossRef]
29. Ahmadianfar, I.; Heidari, A.A.; Gandomi, A.H.; Chu, X.; Chen, H. RUN beyond the Metaphor: An Efficient Optimization Algorithm Based on Runge Kutta Method. *Expert Syst. Appl.* **2021**, *181*, 115079. [CrossRef]
30. UCI Machine Learning Repository. Adult. 1996. Available online: https://archive.ics.uci.edu/dataset/2/adult (accessed on 4 March 2026).
31. UCI Machine Learning Repository. Bank Marketing. 2014. Available online: https://archive.ics.uci.edu/dataset/222/bank+marketing (accessed on 4 March 2026).
32. UCI Machine Learning Repository. Default of Credit Card Clients. 2009. Available online: https://archive.ics.uci.edu/dataset/350/default+of+credit+card+clients (accessed on 4 March 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.