Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 2, April 2026, pp. 1756~1769
ISSN: 2302-9285, DOI: 10.11591/eei.v15i2.11224

Anonymization techniques for privacy-preserving data publishing: a comprehensive survey

Sami Smadi[1], Nader Abdel Karim[2], Hasan Kanaker[3], Waleed K. Abdulraheem[2]
[1]Department of Information Technology, Faculty of Information Technology and Computer Sciences, Yarmouk University, Irbid, Jordan
[2]Department of Intelligent Systems, Faculty of Artificial Intelligence, Al-Balqa Applied University, Al-Salt, Jordan
[3]Department of Information Security, Faculty of Information Technology, University of Petra, Amman, Jordan

Article Info

Article history:

Received Aug 17, 2025
Revised Jan 24, 2026
Accepted Mar 5, 2026

Keywords:

Anonymization
Differential privacy
Federated learning
K-anonymity
l-diversity
Pseudonymization
t-closeness

ABSTRACT

Data-driven innovation in healthcare, finance, and smart cities increasingly depends on sharing rich datasets, but such sharing raises severe privacy risks and regulatory challenges. Privacy-preserving data publishing (PPDP) seeks to release useful data while preventing re-identification and inference attacks. This paper presents a comprehensive survey of anonymization techniques for PPDP, spanning traditional models (k-anonymity, l-diversity, t-closeness, and pseudonymization) and modern approaches (differential privacy (DP), synthetic data generation, federated learning (FL), secure multi-party computation (SMPC), homomorphic encryption (HE), blockchain-based schemes, and quantum-safe cryptography). We propose a taxonomy that organizes these methods by privacy guarantees, data utility, scalability, and computational cost, and we provide a comparative analysis of their strengths, limitations, and typical application domains. The survey also reviews legal and ethical frameworks, with particular attention to general data protection regulation GDPR, health insurance portability and accountability act (HIPAA), and related regulations, and highlights emerging trends such as artificial intelligence (AI-driven) anonymization and privacy risks from large language models (LLMs) and quantum computing. Overall, the study shows that various techniques fail to protect all data scenarios so we need to create hybrid systems which will provide explainable anonymization solutions at different scales to protect privacy and maintain useful data utility.

Corresponding Author:

Sami Smadi
Department of Information Technology, Faculty of Information Technology and Computer Sciences
Yarmouk University
Irbid, 21163, Jordan
Email: sami.smadi@yu.edu.jo

1. INTRODUCTION

The research team employed systematic survey methods through four stages to achieve study transparency and reproducibility. The research process consisted of four stages beginning with planning then literature search followed by screening and ending with synthesis.

We first defined the scope of the review as anonymization techniques for privacy-preserving data publishing (PPDP), including both traditional and modern approaches together with their legal and ethical context. We then searched major digital libraries—IEEE Xplore, ACM Digital Library, Scopus, Web of Science, ScienceDirect, SpringerLink, and PubMed—using combinations of the following keywords and phrases: “privacy-preserving data publishing,” “data anonymization,” “microdata de-identification,” “k-

anonymity," "l-diversity," "t-closeness," "differential privacy," "synthetic data," "federated learning," "secure multi-party computation," "homomorphic encryption," "blockchain privacy," and "quantum-safe cryptography." The research examined publications which appeared between 2002 and May 2025 with special focus on studies from 2019 and later.

The research includes peer-reviewed journal articles and full conference papers which fulfill three conditions: i) the papers either present new anonymization methods or PPDP techniques or they analyze these methods or evaluate their effectiveness, ii) the papers conduct surveys and create taxonomies and perform comparative analyses of privacy-preserving mechanisms which affect data publishing operations, and iii) the papers examine both regulatory and ethical issues which affect the privacy protection of released datasets through anonymization methods. We excluded short workshop papers, non–peer-reviewed reports, patents, purely theoretical cryptographic results without a data publishing scenario, and papers focusing solely on access control or secure communication without anonymization.

The search results produced 450 distinct records which became 312 after removing duplicates. Two authors conducted title and abstract screenings to remove studies that did not meet criteria before they conducted full-text evaluations of the remaining papers. Disagreements were resolved through discussion. The research produced 55 primary studies which included 38 journal articles and 17 conference papers. The research team used backward and forward snowballing techniques to grow their initial dataset through the identification of vital surveys and highly referenced papers which presented connected information although they lacked the specific search keywords.

We retrieved the following details from each study we chose: i) the study used which anonymization model or mechanism; ii) the study focused on which specific data type and what particular application field; iii) the study based its threat assessment on which privacy protection standards; iv) the study assessed how well the data maintained its usefulness while evaluating its ability to scale and its computational requirements; v) the study provided complete information about its implementation process along with available software tools; and vi) the study recognized any legal or moral obligations which needed to be addressed. The information helped us create Figure 1 and Table 1 content and establish the structure for discussing privacy utility tradeoffs and legal frameworks and remaining research questions.

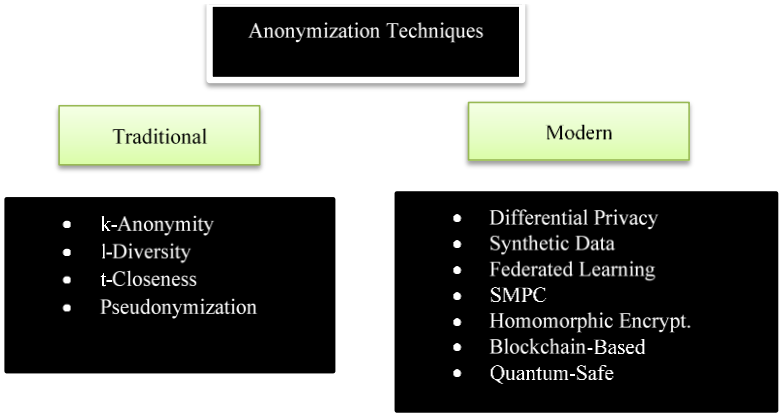

Figure 1. Taxonomy of traditional and modern anonymization techniques for privacy-preserving data publishing

The digital era has made data the fundamental force which drives healthcare, financial, transportation, and smart city innovations [1]. Organizations now have access to extensive data collections because connected devices link to social media platforms and internet of things (IoT) ecosystems which enable predictive analytics and personalized service delivery [2]. The process of data anonymization becomes vulnerable to identification through high-profile incidents which include the Netflix Prize re-identification and the Cambridge Analytica scandal [3]-[5]. The new data sharing methods have created essential barriers which need to protect personal data while maintaining confidentiality standards and public trust.

Problem definition PPDP aims to release useful datasets while preventing re-identification, attribute inference, and membership attacks [6]. In practice, organizations must balance three conflicting objectives: privacy protection, analytical utility, and scalability [7]. Overly aggressive anonymization or perturbation can destroy statistical structure and model performance, whereas weak protection exposes individuals to harms such as discrimination, reputational damage, and regulatory penalties [8]. Regulations including the general data

protection regulation (GDPR), the California consumer privacy act (CCPA), and the health insurance portability and accountability act (HIPAA) further require organizations to justify anonymization choices, document residual risks, and demonstrate compliance when sharing data for research or operational analytics [9].

Table 1. Comparative analysis of major anonymization techniques in PPDP scenarios

Technique	Privacy guarantee	Data utility	Scalability	Computational complexity	Regulatory compliance
k-anonymity	Weak	High	Moderate	Low	Limited
l-diversity/t-closeness	Moderate	Moderate	Moderate	Moderate	Moderate
DP	Strong	Variable	High	Moderate	High
Synthetic data	Moderate-strong	High	High	High	High
FL	Strong	Variable	High	High	High
SMPC	Strong	Variable	Low	Very high	High
HE	Strong	High	Low	Very high	High
Blockchain-based	Moderate-strong	Variable	Low	High	Moderate
Quantum-safe	Strong	Variable	Moderate	High	High

A wide range of anonymization techniques has been proposed for PPDP [10], [11]. Classical models such as k-anonymity, l-diversity, and t-closeness remain influential, and recent work continues to refine them through optimization-based and clustering-based algorithms that reduce information loss and mitigate skewness and homogeneity attacks [12]. In parallel, modern approaches—including DP, synthetic data generation, FL, SMPC, HE, and blockchain-based access control—have been adopted in domains such as healthcare, finance, and smart cities [13], [14]. Several surveys review subsets of this landscape, focusing on big data anonymization, microdata de-identification, DP, or synthetic data, but they typically concentrate on specific techniques or sectors rather than providing a unified perspective.

Despite this rich literature, there is still limited work that jointly: i) compares traditional and modern anonymization techniques under a unified taxonomy, ii) analyzes privacy–utility–scalability trade-offs across heterogeneous application domains, iii) connects technical mechanisms to concrete legal and ethical requirements, and iv) discusses emerging threats from AI-driven attacks and quantum computing together with corresponding quantum-safe and hybrid solutions. Existing surveys often address a single family of methods or a single domain and rarely integrate attack models, regulatory guidance, and deployment considerations into one coherent framework.

The survey evaluates all anonymization methods which defend PPDP data from disclosure. The research contains three essential findings which are: i) systematic taxonomy: the study establishes a classification system which organizes anonymization methods into two categories of traditional and modern techniques that include quantum-safe and blockchain-based solutions while showing their relationship to privacy threats and attack models (Figure 1), ii) in-depth comparative analysis: the research performs an extensive evaluation of privacy protection systems and data utility and system operational performance and calculation needs and official regulations for different anonymization techniques which are shown in Table 1, iii) emerging trends: the research identifies three new trends which include artificial intelligence (AI-based) data protection methods and privacy threats from big language models and generative AI systems and decentralized systems based on blockchain technology and FL and quantum-resistant encryption methods, iv) legal and ethical context: the research investigates how GDPR and HIPAA and CCPA and other privacy regulations affect anonymization methods while it identifies ethical problems which stem from privacy protection data distribution methods, v) future directions: the paper presents current research obstacles which need solution to develop deployable PPDP pipeline-based anonymization systems that provide scalability and robustness and explainable results, and vi) significance: the survey combines attack models with current and previous anonymity techniques and legal frameworks and ethical standards and modern technological advancements to establish a unified resource which serves users from academic and professional backgrounds. The system enables users to choose methods for data protection while performing risk evaluations and designing systems which protect sensitive information during healthcare and financial and smart city data publication. The research develops advanced hybrid and explainable anonymization systems which achieve improved results in privacy protection and data utility and system performance.

The subsequent sections of this survey are organized as follows. Section 2 conducts a detailed examination of privacy threats along with their attack models. Section 3 assesses the effectiveness of conventional data protection methods which use anonymization techniques. Section 4 reviews modern techniques which improve existing data protection systems. Section 5 provides a detailed evaluation of the privacy preservation versus data utility trade-offs. The legal and ethical frameworks which control privacy practices are explained in section 6. The current obstacles are described in section 7 while potential research directions are specified. The section 8 of the paper combines essential findings with concluding statements.

2. PRIVACY THREATS AND ATTACK MODELS

The process of anonymization needs to be based on the understanding of adversary capabilities. This section explains privacy threats, attack models, and their implications, with real-world examples and emerging risks.

2.1. Re-identification attacks

The re-identification attacks reveal individual identities by connecting anonymized records to external datasets. The linkage attacks involve matching quasi-identifiers (e.g., age, gender, and zone improvement plan (ZIP) code) with external sources. Narayanan and Shmatikov [15] demonstrated how Netflix user de-anonymization occurs through matching movie ratings with internet movie database (IMDb) profiles even when data is sparse. The re-identification of medical records became possible through the use of voter registration data according to Sweeney [16]. The singling-out attacks involves separating a record into a small group which makes re-identification more probable [17]. A person becomes identifiable through unique attribute combinations even when there are no exact matches in the dataset. The temporal re-identification attack emerges when dynamic datasets like mobility traces allow adversaries to track records across time through behavioral pattern analysis [18].

2.2. Background knowledge attacks

The attackers can use their existing knowledge to identify important personal details which belong to the victim. For instance, if an individual's demographic information is available and it is known that it is part of a dataset, it is possible to deduce confidential information such as medical conditions. Such inference attacks typically rely on two main sources of auxiliary information. Researchers can identify political affiliations through survey data which seems anonymous by analyzing external knowledge from public records and social media content. Second insider threats create a major security threat because people who have authorized access to specific parts of the dataset including data analysts can purposefully or accidentally violate privacy standards. The scenarios demonstrate why organizations need to implement strong anonymization methods which protect data from being traced back to individuals and stop sensitive information from escaping [19].

2.3. Homogeneity and skewness attacks

Traditional anonymization techniques face essential security risks because of homogeneity and skewness attacks which target k-anonymity. A homogeneity attack occurs when all records in an equivalence class possess identical sensitive attributes which allows an adversary to determine these attributes with absolute certainty [20], [21]. The hospital dataset becomes vulnerable to privacy breaches when patients are grouped by ZIP code because all members of a particular group share the same medical diagnosis. The skewness attack takes advantage of unbalanced distributions of sensitive attributes [22]. Attackers can make probable inferences about rare attribute occurrences through skewness attacks because these attributes dominate their respective equivalence classes. The attacks demonstrate that k-anonymity fails to safeguard data so researchers must create advanced privacy models which will serve as defense mechanisms.

2.4. Inference attacks in machine learning

Machine learning integration into data-driven systems creates new privacy vulnerabilities which attackers can use to their advantage.

- The model inversion attacks: which allow attackers to rebuild original training data from the output of models. The research by Fredrikson *et al*. (2015) [23] proved that neural networks trained on images could reveal personal identities through facial feature extraction.
- The membership inference attacks: the main goal of membership inference attacks attack is to identify which data points were used to train models while creating substantial privacy risks in healthcare and other sensitive domains through FL. Shokri *et al*. (2017) [24] demonstrated how healthcare models could be targeted to infer the participation of individual patients.
- The membership inference attacks: the main goal of membership inference attacks attack is to identify which data points were used to train models while creating substantial privacy risks in healthcare and other sensitive domains through FL. Shokri *et al*. (2017) [24] showed that healthcare systems could be used to identify which patients took part in the study.
- The attribute inference attacks: which try to discover private characteristics such as ethnicity or medical conditions through non-sensitive input data. The study by Yeom *et al*. (2018) [25] demonstrated that genomic models would reveal hidden characteristics by accident.
- The adversarial attacks: is to create specific inputs which take advantage of model vulnerabilities to reveal private information or modify model responses. The research by Xu *et al*. (2023) [26] demonstrates

how attackers can use these inputs to modify model outputs while extracting confidential data. These attack vectors demonstrate the urgent requirement for strong privacy-protecting solutions in machine learning systems.

2.5. Emerging threats

The domain of data privacy experiences increased security threats because attackers implement sophisticated methods and emerging technologies to execute their attacks. Graph-based attacks use the structural features of graph-based data including social networks to identify individuals after anonymization processes. The research by Backstrom *et al*. (2007) [27] showed how attackers use network connections between nodes to overcome privacy protection systems. The emergence of large language models (LLMs) has created a new threat through AI-driven attacks which can reveal sensitive information from text data that appears to be anonymous. The research by Carlini *et al*. [28] revealed that these models can retrieve private content from training data which requires the creation of new defensive system. Quantum attacks will become the future security threat which will compromise all current cryptographic systems that protect anonymity. The encryption systems which Shor's algorithm [29] can decrypt prove that privacy-preserving technologies need quantum-resistant solutions. The new vectors demonstrate why organizations need to predict and solve privacy risks which impact current and future computational systems.

3. TRADITIONAL ANONYMIZATION TECHNIQUES

Traditional methods group individuals to obscure identities, focusing on quasi-identifiers and sensitive attributes. This section expands on their mechanisms, applications, and limitations.

3.1. k-anonymity

The privacy model k-anonymity protects each unique combination of quasi-identifiers by ensuring they exist in at least k records so individual records become indistinguishable when grouped with k or more records [16]. Multiple essential systems need to operate correctly for this principle to become effective. The generalization technique replaces detailed attribute values with more general categories through the example of converting 32 years old into the age bracket "30–39". The suppression method removes both specific data points and complete records which present re-identification threats because they exist in infrequent ZIP code regions. The bucketization process uses pre-defined bins to convert numerical data while maintaining order relationships but it decreases the accuracy of the information.

The technique has become a required method for releasing public datasets which includes census microdata and healthcare records anonymization to fulfill privacy regulations. The deployment of k-anonymity faces various operational barriers which affect its implementation process. The k-anonymity model remains vulnerable to homogeneity attacks which occur when all records in a group have identical sensitive attributes and background knowledge attacks that exploit external information to break anonymity. The process of increasing K values typically requires excessive generalization which leads to substantial data utility reduction.

The Mondrian algorithm [30] improves k-anonymity protection for multiple-dimensional data sets by using an optimization method which maintains data utility while providing enhanced privacy protection. Research has developed k-anonymity through two optimized methods which combine optimization techniques with clustering algorithms to achieve enhanced privacy-data disclosure tradeoffs while protecting against attacks which use similarity patterns and data distribution imbalances in complex datasets. Complex data environments require advanced anonymization models because these systems need to solve basic problems which protect data while operating system functions.

3.2. l-diversity

The privacy model l-diversity protects sensitive information by requiring each equivalence class in a dataset to contain at least l different sensitive attribute value [21]. The model solves specific vulnerabilities in k-anonymity through its introduction of semantic diversity in sensitive attributes which minimizes attribute disclosure risks.

The privacy model l-diversity has received multiple extensions to fulfill different privacy requirements. The privacy model Distinct l-diversity requires each equivalence class to contain l different sensitive values, but entropy l-diversity uses entropy maximization to defend against inference attacks. The privacy model recursive (c, l)-diversity adds a supplementary restriction which limits the prevalence of dominant sensitive values to protect against attribute disclosure risks.

The implementation of l-diversity proves useful in medical records and survey data because it enhances k-anonymity by requiring diverse diagnoses or survey responses among similar demographic

groups. The model contains several drawbacks in its implementation. The model remains susceptible to skewness attacks because of its weakness against uneven distributions of sensitive attributes while its utility decreases when diversity constraints become too strict because it leads to excessive data generalization.

For example, the hospital dataset must contain a minimum number of different diagnoses for each demographic category under l-diversity enforcement. The dominance of the common flu in the dataset could potentially expose sensitive information even when diverse requirements are in place.

3.3. t-closeness

The privacy model t-closeness controls how sensitive attributes within any equivalence class can differ from their overall dataset distribution through a predefined threshold t which uses statistical distance metrics like Earth Mover's distance [31]. The distributional similarity constraint protects sensitive attributes from disclosure risks because it maintains minimal disclosure risks.

The main goal of t-closeness is to stop attackers from exploiting statistical differences in local data subsets through skewness and inference attacks. The model protects against privacy breaches by maintaining global distribution characteristics in each equivalence class.

The privacy model t-closeness functions best with numerical sensitive attributes in datasets which contain income levels and medical test results because it preserves statistical properties to protect against attribute inference attacks. The model contains specific difficulties which need to be addressed. The requirement for exact distributional matching leads to substantial data utility deterioration which becomes more severe when working with restricted datasets that exhibit limited natural data variation. The calculation of distributional distances becomes difficult when working with big datasets that have many dimensions because of performance challenges that arise from complex computations. The t-closeness framework has developed through various modifications which employ Kullback-Leibler (KL) divergence as a distance metric to support different data characteristics and analytical requirements.

3.4. Pseudonymization

The pseudonymization process requires the substitution of personal information and social security numbers with pseudonymous tokens which function as coded identifiers that maintain direct identification capabilities. The substitution process makes it harder to identify data directly but the information stays traceable because of the linkage keys that exist in the data. The GDPR treats pseudonymized data as personal information which must follow all regulatory standards while the article 29 working party [32] provides supervisory guidance that pseudonymization reduces data identifiability but does not eliminate it so organizations need to implement appropriate technical and organizational security measures.

The banking industry together with research organizations protect privacy through pseudonymization because this method enables them to use their analytical systems while maintaining data protection. The research method contains a major weakness because pseudonymized data becomes vulnerable to identification threats when researchers combine additional datasets or conduct linkage or perform inferential attacks. The process of pseudonymization functions independently from true anonymization because it does not provide a full protection against revealing personal identities. The analysis of clinical-trial datasets becomes private through coded IDs instead of patient identifiers but external records and registries could reveal the true identities of patients.

3.5. Practical considerations

The implementation of anonymization techniques requires domain experts to conduct complete assessments for quasi-identifier selection because this process requires their specialized knowledge. The risks from insufficient anonymization are dual because both excessive generalization or suppression result in major data utility loss and insufficient anonymization makes privacy breaches through re-identification more likely. The software tools anonymization and re-identification risk analysis (ARX) and Amnesia offer specific frameworks which let users run established anonymization models such as k-anonymity and l-diversity to solve these problems. The tools which Prasser *et al.* [33] describe how to enable users to set up anonymization parameter settings through interfaces which help them assess and enhance protection levels without compromising data analysis capabilities. Research conducted during recent times combines these tools with improved k-anonymity and l-diversity algorithms and clustering methods which adjust to particular domains to obtain superior privacy-utility tradeoffs for modern high-dimensional and streaming data sets.

4. MODERN ANONYMIZATION APPROACHES

Modern techniques solve existing field problems through mathematical methods which combine cryptographic approaches and machine learning algorithms. The following section explains the operational systems and their current state of development and actual system implementations.

4.1. Differential privacy

DP is a formal privacy framework that ensures the inclusion or exclusion of any individual's data in a dataset has a negligible impact on the outcome of an analysis. A mechanism satisfies (ε,δ) DP requirements when it modifies individual data records in ways that produce output probability changes which stay within the established ε (privacy budget) and δ (failure probability) limits [34]. The privacy-utility trade-off becomes quantifiable through ε which shows the extent of privacy information disclosure.

Multiple systems exist to put DP into operational use. The Laplace mechanism adds Laplace distribution noise to numerical query results which depends on the query sensitivity level. The Gaussian mechanism achieves DP through Gaussian noise addition which delivers the best results for systems that need to run multiple iterations. The exponential mechanism enables DP to handle non-numerical outputs through a utility function which determines the selection of results through probabilistic methods.

DP has been adopted in large-scale, real-world applications. The U.S. Census Bureau used DP to defend demographic information during the 2020 Census [35]. The randomized aggregatable privacy-preserving ordinal response (RAPPOR) framework from Google uses local perturbation to collect telemetry data which becomes part of the framework [36] while Apple implements DP to study combined emoji usage statistics without exposing user information. The field has seen two major developments which involve local DP for pre-aggregation data perturbation and privacy-enhancing FL that combines DP with decentralized model training for better security and scalability [37].

The system DP keeps its useful characteristics but faces multiple major operational problems which affect its performance. The introduction of noise leads to major precision deterioration which becomes most difficult when working with restricted data sets and complex database queries. The selection of an appropriate ε value becomes difficult because it depends on particular situations. OpenDP and TensorFlow Privacy operate as open-source libraries which enable researchers and practitioners to perform DP data analysis through their ability to preserve privacy while maintaining useful results.

4.2. Synthetic data generation

The process of synthetic data generation creates artificial datasets which maintain both statistical properties and structural elements of actual data through the implementation of generative adversarial networks (GANs) [38] and variational autoencoders (VAEs) and diffusion models [39]. The methods become more secure when DP protects them during training because they reduce the chances of identifying individual data points [38].

This method has become a standard tool which scientists use across different fields of study. Synthetic electronic health records (EHRs) in healthcare enable research and model development without compromising patient privacy [39]. The financial industry uses synthetic transaction datasets to fight fraud and build risk models and fulfill regulatory requirements. Social science researchers protect confidential survey answers through synthetic datasets which allows them to share data freely while they verify their research results. The latest advancements in diffusion-based generative models have achieved better results than GANs for producing authentic tabular data which maintains its original statistical properties [40].

The development of synthetic data requires solutions of various obstacles which need to be resolved during its initial creation stage. The search for proper data privacy protection methods against real-world data becomes difficult because untrained models generate artificial data which lacks realistic characteristics and displays biased patterns while real-world data might accidentally disclose personal information. The process of creating high-quality synthetic datasets which operate in complex high-dimensional domains becomes too expensive to afford. The Synthpop framework demonstrates its operational value through its ability to create fake census data which maintains both statistical connections and data distribution patterns for researchers who require privacy protection of personal information [41].

4.3. Federated learning

FL represents a distributed machine learning system which allows users and institutions to train their local models from their data while sending model updates to a central server for coordination. The architecture maintains data decentralization which minimizes the need to expose raw datasets directly [42].

FL has been applied in diverse contexts. The predictive model of Gboard from Google uses FL to enhance its next-word prediction system while protecting user input data from being stored. The healthcare sector uses FL to create diagnostic models which multiple hospitals can work on together while patients maintain their privacy [43].

FL provides several benefits, but it encounters multiple significant obstacles. The system continues to face privacy threats because attackers can use membership inference attacks to determine which data points were used for model training. The model convergence rate decreases and global model performance suffers when clients have different data distributions that are not independent from each other. The

development of personalized FL solves these problems through model adaptation which generates global models that learn client data patterns to achieve improved accuracy and customized results [44].

4.4. Secure multi-party computation

The SMPC cryptographic system allows different parties to run joint computations on their confidential data while keeping their input information hidden from all other participants. The three primary methods of SMPC involve secret sharing which splits data into encrypted shares distributed among participants and garbled circuits which transform computations into encrypted Boolean circuits and HE which enables computations on encrypted data [45].

The SMPC protocol allows privacy-sensitive domains to function through its deployment in financial services for credit scoring and healthcare research collaborations which protect patient information. The practical use of this technology has become more feasible through recent developments which include the SPDZ protocol family that provides enhanced performance through its precomputation optimization and secure arithmetic operation capabilities [46].

The organization continues to encounter ongoing difficulties. The system encounters difficulties in its growth because it maintains high operational costs for processing and data transfer when dealing with large or urgent data collections. The security of certain protocols becomes vulnerable to trust-related issues because they require trusted setup phases for their operation. The Sharemind platform demonstrates the practical application of SMPC through its deployment which allows organizations to securely analyze joint data by using secret sharing methods to defend their sensitive information while performing reliable statistical operations [47].

4.5. Homomorphic encryption

HE operates as a cryptographic system which allows secure operations on encrypted data without decryption needs thus maintaining complete data privacy throughout the entire process. FHE allows unrestricted computations on ciphertexts, but PHE operates with restricted operations that depend on the chosen scheme [48]. HE has been applied in a range of privacy-critical domains. Secure cloud computing enables service providers to handle protected client information through encryption which prevents them from obtaining access to the unencrypted data. Private machine learning operations become possible through HE because it allows model training and inference operations on encrypted data which maintains dataset privacy during outsourcing and collaborative work [49].

The practicality of HE has become more functional because of recent developments which benefit machine learning operations. The CKKS scheme enables approximate arithmetic operations which provide an efficient framework for encrypted computations of real numbers that are prevalent in ML tasks [50]. The fully homomorphic form of HE requires high computational power which generates substantial performance and memory consumption that makes it difficult to deploy in big systems and time-sensitive applications. The open-source libraries Microsoft SEAL and HElib enable adoption through their optimized implementations which let researchers and practitioners test and implement HE for privacy protection in systems.

4.6. Blockchain-based anonymization

The blockchain-based anonymization system protects privacy through distributed ledger technology which maintains privacy while providing transparency through its built-in accountability features. The tamper-resistant audit trails maintain permanent records of anonymization processes which enable stakeholders to verify compliance while building trust. Organizations can prove their privacy and security policy compliance through zero-knowledge proofs (ZKPs) which verify compliance without revealing any sensitive information to protect confidentiality standards and meet regulatory needs [51].

The system allows healthcare organizations to exchange medical information through secure channels which protect patient confidentiality and supply chain operators to show their operations without exposing their confidential business information. The implementation of zero-knowledge succinct non-interactive arguments of knowledge (zk-SNARKs) has brought major improvements to ZKP-based systems by decreasing proof generation time [52].

The deployment of blockchain-based anonymization systems encounters various technical barriers which include performance restrictions and large data storage requirements and proof-of-work systems that use substantial amounts of electricity. The expenses generate opposition against the strong audit capabilities and tamper-proof features which distributed ledgers provide. Public blockchains function without authorization yet they enable users to see everything and defend against censorship attempts yet their performance issues with transaction speed and high fees prevent them from supporting big PPDP operations. Permissioned blockchains achieve lower latency and reduced energy consumption through their implementation of light-weight consensus protocols yet they compromise on network decentralization and create potential governance challenges. The MedRec system shows how blockchain-based audit trails create

trust for electronic medical record sharing but it requires specific architectural decisions to achieve both auditability and performance and sustainability goals [53].

4.7. Quantum-safe anonymization

Quantum-safe anonymization protects encrypted datasets from quantum computer decryption through lattice-based encryption and other post-quantum cryptographic methods [29]. The integration of quantum-resistant algorithms into PPDP processes provides enduring confidentiality protection for longitudinal datasets with extended sensitivity periods such as genomic records and national security archives.

The U.S. National Institute of Standards and Technology (NIST) has standardized post-quantum cryptographic algorithms with Kyber being one of the schemes that approaches formal deployment [54]. These developments work to establish strong security measures for the expected quantum computing environment.

The implementation of quantum-safe anonymization methods presents several significant obstacles. The computational requirements of post-quantum cryptographic methods exceed those of traditional cryptographic methods which can negatively affect system performance in extensive applications. The implementation of quantum-safe infrastructure demands substantial modifications to present systems and communication protocols. The implementation of quantum-safe digital signatures for blockchain-based PPDP frameworks serves as a practical example according to Fernandez-Carames and Fraga-Lamas [55] to protect against both classical and quantum adversaries.

5. UTILITY-PRIVACY TRADE-OFFS

The primary difficulty of PPDP requires finding the right equilibrium between data usefulness and privacy protection because enhanced privacy protection methods reduce the analytical worth of the data. Organizations need to assess their numerical performance indicators through business requirement assessment and rigorous testing procedures to achieve their best possible equilibrium.

5.1. Metrics for data utility

The evaluation process for PPDP methods needs robust quantitative assessment methods to determine the extent of analytical value preservation in anonymized or perturbed datasets. Information loss metrics assess the extent of data distortion which occurs during transformation procedures by using KL divergence and mean squared error and generalization loss to measure data distribution changes from their original state [19]. The evaluation process for essential statistical characteristics in data maintenance uses moment comparison methods to check mean and variance values and distributional equivalence through the Kolmogorov–Smirnov test. The dataset proves its value for particular applications through performance metrics which assess its capabilities by using classification accuracy and regression R2 scores and clustering quality indices. The evaluation of query accuracy helps determine how well aggregate query results match their original values after adding noise for DP protection.

5.2. Quantitative evaluation

The impact of PPDP methods on data utility depends on the privacy model and its parameterization. k-anonymity maintains high utility for basic aggregate queries yet leads to substantial utility loss in advanced analytics because it requires quasi-identifier generalization and suppression [30]. The privacy budget (ε) determines the trade-off between privacy guarantees of DP because smaller values of ε enhance privacy protection at the expense of accuracy according to the U.S. Census Bureau's 2020 application which struggled with accuracy in small geographic areas [35]. Synthetic data generation maintains high utility when generative models accurately model variable relationships and correlations yet overfitting especially in GANs reduces generalizability and affects downstream performance [39]. The performance of FL heavily relies on client data heterogeneity because DP noise addition typically leads to 5–10% predictive accuracy reduction in specific applications [44].

5.3. Application-specific considerations

The relationship between privacy and utility in PPDP depends on specific domain requirements which include operational needs and regulatory requirements and analytical needs. The healthcare field requires high utility for diagnosis prediction and treatment outcome modeling but must maintain strict privacy protection to meet regulatory standards and protect patient information. The combination of synthetic data generation with DP techniques enables collaborative research by protecting identifiable data according to Rieke *et al*. [43]. The financial sector depends on fraud detection systems to detect rare patterns in

transaction data for accurate identification purposes. HE enables analysts to perform operations on encrypted data but its high processing requirements prevent its use in real-time and big data applications [49] .The high-speed and large-scale nature of IoT data streams requires efficient and scalable anonymization methods because they produce continuous data streams. The FL system operates correctly in restricted distributed networks because it enables edge devices to develop models independently without requiring them to transmit their complete data. The combination of bandwidth constraints with device variability creates major obstacles which prevent large-scale IoT network deployments from achieving their maximum model performance [44].

5.4. Comparative analysis

The evaluation of prominent anonymization techniques in Table 1 shows their performance based on privacy guarantees, data utility, scalability, computational complexity, and regulatory compliance. The evaluation in Table 1 provides a general comparison between anonymization methods while Figure 1 presents their conceptual classification structure.

The evaluation process shows that privacy strength and practical deployability represent two opposing factors which appear in all assessment results. The combination of k-anonymity provides high data utility but low computational cost which results in weak privacy protection that fails to meet strict regulatory standards. The privacy level of l-diversity and t-closeness methods exists at a moderate level because these improvements reduce data usefulness and extend processing duration.

The mathematical privacy protection of DP depends on the privacy budget (ε) which requires exact parameter settings to maintain its analytical strength. The process of creating synthetic data produces useful results which maintain acceptable privacy standards according to formal privacy models but it requires substantial computational resources. Organizations can run collaborative analytics through FL and SMPC which protects privacy by distributing data across different locations. The performance of FL depends on the diversity of its data distribution yet SMPC encounters significant scalability problems because of its expensive computation and data exchange requirements. HE provides both strong privacy safeguards and useful data operations on encrypted data but its complex computation makes it unsuitable for big-scale or urgent applications. The two main obstacles which prevent organizations from deploying these solutions at large scale stem from blockchain-based systems with quantum-safe anonymization methods which provide strong privacy protection yet introduce performance delays and make integration more difficult and affect system scalability.

The research evaluates privacy protection functions for data publishing and anonymization through an assessment approach which differs from previous investigations about these domains. Existing surveys often concentrate on big data anonymization, specific mechanisms such as DP or synthetic data, or particular application domains such as healthcare or scientific data. The research adds new information to existing knowledge through Table 1 which combines traditional and contemporary methods to create a single assessment system that assesses both regulatory compliance and business growth opportunities. The framework provides healthcare and financial and smart city practitioners with methods to choose techniques which fulfill both technological requirements and legal standards while keeping enough analytical power for risk prediction and fraud detection and policy assessment.

6. LEGAL AND ETHICAL CONSIDERATIONS

The process of anonymization requires technical design expertise and legal compliance and ethical responsibility to achieve proper privacy protection and data utility balance. The GDPR classifies pseudonymized data as personal data which requires strong protection measures, but truly anonymized data receives exemption status although proving irreversibility remains difficult because of auxiliary data risks. Organizations that fail to comply with regulations face substantial penalties such as Google received a €50M fine in 2019.

HIPAA in the United States provides two de-identification methods through Safe Harbor which requires the removal of 18 identifiers and expert determination which needs statistical proof of low re-identification risk. These frameworks support valuable applications through de-identified EHRs in healthcare research but create operational challenges because Safe Harbor lacks flexibility and expert determination need specialized expertise. The CCPA (California), APPI (Japan), and LGPD (Brazil) together with GDPR principles have started to establish anonymization as a fundamental component of data governance systems across their jurisdictions.

The ethical implications of anonymization have become essential issues for public discussion. The improper implementation of design systems results in biased outcomes which produces unfair treatment of minority groups as studies have shown that credit scoring systems produce discriminatory results when using anonymized data. Organizations need to maintain public trust by explaining their methods transparently and

getting proper consent from vulnerable groups and setting up clear technical oversight through DP's ε value. Strong anonymization practices need to fulfill legal requirements and preserve social equality to defend privacy rights which should protect all communities with equal protection.

7. OPEN CHALLENGES AND FUTURE DIRECTIONS

The field of PPDP faces various technical and practical and interdisciplinary challenges which slow down its development despite its significant advancements. Real-time processing needs require lightweight solutions which include streaming DP and approximate k-anonymity because these methods work with high-dimensional data and streaming and IoT-generated data while maintaining scalability. The protection of data from adversarial attacks continues to be crucial because AI-based re-identification systems have outperformed conventional security systems and researchers have created effective robust DP and certified defense methods which require additional work for improvement. The complex nature of GANs and fully HE makes it difficult to achieve explainable anonymization because their uninterpretable models create challenges for trust-based systems and regulatory compliance and auditability processes. The growing use of hybrid methods which unite DP formal guarantees with synthetic data utility and FL decentralization requires optimized management to prevent performance degradation. The need for quantum-safe anonymization methods has become essential because quantum computing poses a threat to cryptographic primitives thus lattice-based encryption standards are currently being integrated into sensitive genomic data protection. The need for AI systems which defend user privacy has grown because big language models and foundation models face data exposure risks when memory storage systems fail or information systems get breached but DP-based fine-tuning and federated LLMs serve as security measures. The technical development of PPDP faces various interdisciplinary challenges which need both regional governance frameworks and user-centered design methods to achieve ethical deployment and improve professional and policy-maker comprehension. The protection of privacy requires continuous collaboration between computer scientists and legal experts and ethicists and industry stakeholders to stay ahead of data innovation developments.

8. CONCLUSION

The survey provides a single perspective about anonymization methods which protect personal information during data publication through both traditional and contemporary privacy preservation techniques. The survey examines both traditional privacy models which include k-anonymity and l-diversity and t closeness and pseudonymization and advanced methods which include DP and synthetic data creation and FL, SMPC, HE, blockchain-based systems, and quantum-safe encryption. We built our research on a systematic evaluation of existing studies which led to the development of a classification system that groups these methods based on their privacy protection levels and their ability to preserve data quality and their performance capabilities and their algorithmic requirements and their compliance with privacy regulations. The assessment of privacy threats and attack models and GDPR and HIPAA regulation evaluation shows that PPDP requires technical design implementation to meet legal and ethical requirements instead of depending on algorithmic anonymization for solution. The review demonstrates that different techniques work best for specific situations because no single method provides complete solutions for every environment. The development of hybrid solutions which unite DP with synthetic data and FL systems shows promise for future applications which must also provide explainable results and auditable trails and resistance to quantum attacks. The future research agenda needs to create evaluation techniques and standardized testing tools and multiple areas of expertise to develop efficient PPDP systems which protect privacy while maintaining essential data worth for ethical data-based development.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the editorial team of the Bulletin of Electrical Engineering and Informatics for their constructive comments, which helped to improve the quality and clarity of this paper.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sami Smadi	✓	✓			✓	✓	✓	✓	✓	✓		✓	✓	
Nader Abdel Karim		✓			✓	✓				✓				
Hasan Kanaker	✓				✓	✓	✓		✓	✓	✓		✓	
Waleed K. Abdulraheem	✓				✓			✓		✓			✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

This paper does not involve any new data creation or analysis, so data availability is not relevant.

REFERENCES

[1] D. K. Das, "Exploring the Symbiotic Relationship between Digital Transformation, Infrastructure, Service Delivery, and Governance for Smart Sustainable Cities," *Smart Cities*, vol. 7, no. 2, pp. 806–835, 2024, doi: 10.3390/smartcities7020034.

[2] M. Paramesha, N. L. Rane, and J. Rane, "Big Data Analytics, Artificial Intelligence, Machine Learning, Internet of Things, and Blockchain for Enhanced Business Intelligence," *Partners Universal Multidisciplinary Research Journal (PUMRJ)*, vol. 1, no. 2, pp. 110–133, 2024, doi: 10.5281/zenodo.12827323.

[3] J. Powar and A. R. Beresford, "SoK : Managing risks of linkage attacks on data privacy," *Proceedings on Privacy Enhancing Technologies*, vol. 2, pp. 97–116, 2023, doi: 10.56553/popets-2023-0043.

[4] J. Hinds, E. J. Williams, and A. N. Joinson, ""It wouldn' t happen to me": Privacy concerns and perspectives following the Cambridge Analytica scandal," *International Journal of Human-Computer Studies*, vol. 143, 2020, doi: 10.1016/j.ijhcs.2020.102498.

[5] N. Weiss-Blatt, "Chapter 2: Big Tech – Big Scandals Available to Purchase," in *The Techlash and Tech Crisis CommunicationAvailable to Purchase*, Emerald Publishing Limited, pp. 37-72, 2021, doi: 10.1108/978-1-80043-085-320211007.

[6] A. Majeed and S. Lee, "Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey," in *IEEE Access*, vol. 9, pp. 8512-8545, 2021, doi: 10.1109/ACCESS.2020.3045700.

[7] R. F. Parks, R. T. Wigand, and P. B. Lowry, "Balancing information privacy and operational utility in healthcare : proposing a privacy impact assessment (PIA) framework ABSTRACT," *European Journal of Information Systems*, vol. 32, no. 6, pp. 1052–1069, 2022, doi: 10.1080/0960085X.2022.2103044.

[8] H. Koyuncu and R. Altaher "Classification of data anonymization techniques," in *Artificial Intelligence for Cyber Security and Industry 4.0*, CRC Press, pp. 57–96, 2025, doi: 10.1201/9781032657264-3.

[9] B. C. Kara, C. Eyupoglu, and O. Karakuş, "(r, k, ε)-Anonymization: Privacy-Preserving Data Publishing Algorithm Based on Multi-Dimensional Outlier Detection, k-Anonymity, and ε-Differential Privacy," in *IEEE Access*, vol. 13, pp. 70422-70435, 2025, doi: 10.1109/ACCESS.2025.3559410.

[10] A. B. Mercedes, R. A. Muñoz, and J. Manuel, "Protecting privacy in the age of big data : exploring data linking methods for quasi-identifier selection," *International Journal of Information Security*, vol. 24, no. 1, pp. 1–14, 2025, doi: 10.1007/s10207-024-00944-7.

[11] D. Mcgraw and K. D. Mandl, "Privacy protections to encourage use of health-relevant digital data in a learning health system," *NPJ Digital Medicine*, vol. 4, no. 2, pp. 1–11, 2021, doi: 10.1038/s41746-020-00362-8.

[12] E.Özmen, A.Cherukuri, and A. Jonnalagadda, "Analyzing effects of privacy models on smart healthcare data," *Annals of Mathematics and Computer Science*, vol. 28, pp. 79–105, 2025, doi: 10.56947/amcs.v28.526.

[13] S. Prabowo *et al.*, "Privacy-Preserving Tools and Technologies : Government Adoption and Challenges," *IEEE Access*, vol. 13, pp. 33904–33934, 2025, doi: 10.1109/ACCESS.2025.3540878.

[14] N. A. Karim *et al.*, "Performance Comparison of Hyper-V and KVM for Cryptographic Tasks in Cloud Computing," *Computers, Materials & Continua*, vol. 78, no. 2, pp. 2023–2045, 2023, doi: 10.32604/cmc.2023.044304.

[15] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *2008 IEEE Symposium on Security and Privacy (sp 2008)*, Oakland, CA, USA, 2008, pp. 111-125, doi: 10.1109/SP.2008.33.

[16] L. Sweeney, "Simple demographics often identify people uniquely," *Health (San Francisco)*, vol. 671, no. 2000, pp. 1–34, 2000.

[17] C. Dwork, A. Smith, T. Steinke, and J. Ullman, "Exposed! a survey of attacks on private data," *Annual Review of Statistics and Its Application*, vol. 4, no. 1, pp. 61–84, 2017, doi: 10.1146/annurev-statistics-060116-054123.

[18] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific Reports* vol. 3, no. 1, p. 1376, 2013, doi: 10.1038/srep01376.

[19] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments," *A ACM Computing Surveys (CSUR)*, vol. 42, no. 4, pp. 1–53, 2010, doi: 10.1145/1749603.1749605.

[20] F. Liu and X. Zhao, "Disclosure risk from homogeneity attack in differentially privately sanitized frequency distribution," *IEEE IEEE Transactions on Dependable and Secure Computing*, vol. 20, no. 5, pp. 3927–3939, 2022, doi: 10.1109/TDSC.2022.3220592.
[21] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery From Data*, vol. 1, no. 1, pp. 3-es, 2007, doi: 10.1145/1217299.12173.
[22] B. Su, J. Huang, K. Miao, Z. Wang, X. Zhang, and Y. Chen, "K-anonymity privacy protection algorithm for multi-dimensional data against skewness and similarity attacks," *Sensors*, vol. 23, no. 3, p. 1554, 2023, doi: 10.3390/s23031554.
[23] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *CCS '15: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 1322–1333, doi: 10.1145/2810103.2813677.
[24] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," in *2017 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, USA, 2017, pp. 3-18, doi: 10.1109/SP.2017.41.
[25] S. Yeom, I. Giacomelli, M. Fredrikson, and S. Jha, "Privacy risk in machine learning: Analyzing the connection to overfitting," in *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, Oxford, UK, 2018, pp. 268-282, doi: 10.1109/CSF.2018.00027.
[26] Z. Xu *et al.*, "Federated Learning of Gboard Language Models with Differential Privacy," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 2023, vol. 5, pp. 629–639, doi: 10.18653/v1/2023.acl-industry.60.
[27] L. Backstrom, C. Dwork, and J. Kleinberg, "Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography," in *Proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 181–190, doi: 10.1145/1242572.1242598.
[28] N. Carlini *et al.*, "Extracting training data from large language models," in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 2633–2650.
[29] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, Santa Fe, NM, USA, 1994, pp. 124-134, doi: 10.1109/SFCS.1994.365700.
[30] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Mondrian multidimensional k-anonymity," in *22nd International Conference on Data Engineering (ICDE'06)*, Atlanta, GA, USA, 2006, pp. 25-25, doi: 10.1109/ICDE.2006.101.
[31] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *2007 IEEE 23rd International Conference on Data Engineering*, Istanbul, Turkey, 2007, pp. 106-115, doi: 10.1109/ICDE.2007.367856.
[32] Article 29 Data Protection Working Party, "Opinion 05/2014 on Anonymisation Techniques," WP216, Apr. 2014. [Online]. Available: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.
[33] F. Prasser, F. Kohlmayer, R. Lautenschläger, and K. A. Kuhn, "Arx-a comprehensive tool for anonymizing biomedical data," in *AMIA Annual Symposium Proceedings*, 2014, p. 984.
[34] C. Dwork, "Differential privacy," in *International Colloquium on Automata, Languages, and Programming*, Springer, 2006, pp. 1–12, doi: 10.1007/11787006_1.
[35] J. M. Abowd, "The US Census Bureau adopts differential privacy," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, p. 2867, 2018, doi: 10.1145/3219819.3226070.
[36] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 2014, pp. 1054–1067, doi: 10.1145/2660267.2660348.
[37] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014, 10.1561/0400000042.
[38] M. Abadi *et al.*, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318, doi: 10.1145/2976749.2978318.
[39] E. Choi, S. Biswal, B. Malin, J. Duke, W. F. Stewart, and J. Sun, "Generating multi-label discrete patient records using generative adversarial networks," in *Machine Learning for Healthcare Conference*, PMLR, 2017, pp. 286–305.
[40] A. Kotelnikov, D. Baranchuk, I. Rubachev, and A. Babenko, "Tabddpm: Modelling tabular data with diffusion models," in *International Conference on Machine Learning*, PMLR, 2023, pp. 17564–17579.
[41] B. Nowok, G. M. Raab, and C. Dibben, "synthpop: Bespoke creation of synthetic data in R," *Journal of Statistical Software*, vol. 74, no.11, pp. 1–26, 2016, doi: 10.18637/jss.v074.i11.
[42] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*, PMLR, 2017, pp. 1273–1282.
[43] N. Rieke *et al.*, "The future of digital health with federated learning," *NPJ Digitial Medical*, vol. 3, no. 1, p. 119, 2020, doi: 10.1038/s41746-020-00323-1.
[44] P. Kairouz and H.B. McMahan, "Advances and open problems in federated learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021, doi:10.1561/2200000083.
[45] A. C.-C. Yao, "How to generate and exchange secrets," in *27th Annual Symposium on Foundations of Computer Science (Sfcs 1986)*, Toronto, ON, Canada, 1986, pp. 162-167, doi: 10.1109/SFCS.1986.25.
[46] I. Damgård, V. Pastro, N. Smart, and S. Zakarias, "Multiparty computation from somewhat homomorphic encryption," in *Annual Cryptology Conference*, Springer, 2012, pp. 643–662, doi: 10.1007/978-3-642-32009-5_38.
[47] D. Bogdanov, S. Laur, and J. Willemson, "Sharemind: A framework for fast privacy-preserving computations," in *European Symposium on Research in Computer Security*, Springer, vol 5283, pp. 192–206, 2008, doi: 10.1007/978-3-540-88313-5_13.
[48] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, 2009, pp. 169–178, doi: 10.1145/1536414.153644.
[49] L. B. Pulido-Gaytan, A. Tchernykh, J. M. Cortés-Mendoza, M. Babenko, and G. Radchenko, "A survey on privacy-preserving machine learning with fully homomorphic encryption," in *Latin American High Performance Computing Conference*, Springer, 2020, pp. 115–129, doi: 10.1007/978-3-030-68035-0_9.
[50] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *International Conference on the Theory and Application of Cryptology and Information Security*, Springer, 2017, pp. 409–437, doi: 10.1007/978-3-319-70694-8_15.
[51] J. Priya and C. Palanisamy, "Novel block chain technique for data privacy and access anonymity in smart healthcare," *Intelligent Automation & Soft Computing*, vol. 35, no. 1, pp. 243–259, 2023, doi: DOI:10.32604/iasc.2023.025719.
[52] J. Groth, "On the size of pairing-based non-interactive arguments," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Springer, 2016, pp. 305–326, 2016, doi: 10.1007/978-3-662-49896-5 11.
[53] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using blockchain for medical data access and permission

management," in *2016 2nd International Conference on Open and Big Data (OBD)*, 2016, pp. 25-30, doi: 10.1109/OBD.2016.11.
[54] S. Li *et al.*, "Post-quantum security: Opportunities and challenges," *Sensors*, vol. 23, no. 21, p. 8744, 2023, doi: 10.3390/s23218744.
[55] T. M. Fernandez-Carames and P. Fraga-Lamas, "A review on the application of blockchain to the next generation of cybersecure industry 4.0 smart factories," *IEEE Access*, vol. 7, pp. 45201–45218, 2019, doi: 10.1109/ACCESS.2019.2908780.

BIOGRAPHIES OF AUTHORS

Sami Smadi is an Assistant Professor in the Department of Information Technology, Faculty of Information Technology and Computer Sciences, Yarmouk University, Jordan. He received his Ph.D. from Northumbria University Newcastle, United Kingdom. His M.Sc. degree in Computer Science from Yarmouk University, Jordan in 2006, and his B.Sc. degree in Computer Science from Hashemite University, Jordan. His main research areas are information security, network security, phishing detection, reinforcement learning, machine learning, and artificial neural network. He can be contacted at email: sami.smadi@yu.edu.jo.

Nader Abdel Karim received his Ph.D. degree in Cybersecurity from UKM University, Malaysia, in 2017. Currently, he is an assistant professor at the Department of Cybersecurity, College of Artificial Intelligence, Al-Balqa Applied University, Jordan. In the areas of user authentication, cyber security, human-computer interaction (HCI), and online learning, he has very strong experience. He has also participated in a number of research projects, including ones on virtual privacy techniques and preferences-based authentication. He can be contacted at email: nader.salameh@bau.edu.jo.

Hasan Kanaker received his Ph.D. degree in Cybersecurity from USIM University, Malaysia, in 2018. Currently, he is an assistant professor at the Department of Information Security, Faculty of Information Technology, University of Petra, Jordan. His research interests include security, intrusion detection, machine learning, malware, deep learning, malware detection, network security, user authentication, cloud computing security, and information security. He can be contacted at email: hasan.kanaker@uop.edu.jo.

Waleed K. Abdulraheem received his Ph.D. degree in Cybersecurity from UPM, University, Malaysia, in 2019. Currently, he is an assistant professor at the Cybersecurity Department, College of Artificial Intelligence, Al-Balqa Applied University, Jordan. In the areas of cryptography, cyber security, human-computer interaction (HCI), and cloud computing, he has very strong experience. He has also participated in a number of research projects, including cryptography, and cloud computing security. He can be contacted at email: w.khalid@bau.edu.jo.