IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

INTELLIGENT PRIVACY PRESERVING DATA ENCRYPTION AND ANONYMIZATION SYSTEM

Sangeetha M.E.[1], Sanjeevi R B[2], Vinoth M[3]

Assistant Professor, CSE & Dhanalakshmi Srinivasan College of Engineering & Technology, India[1]

CSE (Cyber Security) & Dhanalakshmi Srinivasan College of Engineering & Technology, India[2]

CSE (Cyber Security) & Dhanalakshmi Srinivasan College of Engineering & Technology, India[3]

Abstract: The project not only addresses privacy and security needs but also the issue of trust in data sharing practices. By providing a privacy-preserving system that guarantees the confidentiality of individuals' information, the project aims at creating a secure and reliable environment for data sharing. Additionally, the proposed system will support the implementation of various use cases such as health care, banking, and e-commerce where data privacy and security are of paramount importance. The integration of various advanced techniques like decision tree, random forest, or support vector machine for data sensitivity assessment and the selection of the most suitable cryptographic technique for the different types of data to be processed will be also feasible. This will help in providing the best possible solution in terms of data privacy and accessibility.Ultimately, the project is expected to make a strong contribution to the development of privacy-preserving technologies through its innovative approach and the development of the proposed Intelligent Privacy Preserving Data Encryption and Anonymization System. In the long run, the project's results might have a significant impact on the future of data protection technologies as more and more organizations migrate to cloud services and share information across borders.

Keywords: Data protection, encryption, anonymization, privacy-preserving, cloud computing, big data analytics, interconnected systems, decision making, automated systems, trust

I. INTRODUCTION

The digital era has given rise to a very rapid and very large data generation and storage which has transformed the way of working of organizations, governments, and individuals. It is worth mentioning that all these transitions have been made possible through the broad use of cloud computing, the Internet of Things (IoT), big data analysis, and artificial intelligence to name just a few, which have resulted in continuous collection, processing, and sharing of sensitive data over various systems. Though, this data-centric environment is a boon to innovation and efficiency, at the same time it is fraught with very serious concerns regarding data privacy, security, and unauthorized access. Protection of sensitive information from cyber threats, data breaches, and misuse has now become one of the topmost challenges confronted by the modern information systems industry.The conventional data protection paradigm mainly revolves around encryption as the primary method for securing data during the phases of storage and transmission. Encryption is the process through which data is kept a secret by converting readable information to an unreadable format through the use of cryptographic keys. But in most cases when, data is shared, analyzed or processed by third parties, encryption alone is not sufficient. Once decrypted, the sensitive information can be exposed, attacked from within or misused. Moreover, with the growing amount of computing power and the emerging cyber-attack techniques, static encryption models have difficulty in providing the flexibility and long-term security required.Besides encryption, data anonymization has emerged as the principal method for safeguarding individual privacy. The objective of the Anonymization techniques is to eliminate or obscure personally identifiable information (PII) from datasets in such a way that the individuals can neither be recognized directly nor indirectly. Data masking, generalization, suppression, and pseudonymization are the most commonly used methods in healthcare, finance, social media, and research applications. Nonetheless, the traditional anonymization techniques often come with the drawback of re-identification risk, and others.

II. LITERATURE REVIEW

Q. Li, Y. Tian, Y. Zhang, L. Shen and J. Guo., (2025) [1] suggests an effective access control system relying on CP-ABE encryption that however mainly covers access policy protection and not complete data privacy. Besides that, no intelligent

DOI: 10.17148/IJARCCE.2026.155126

or adaptable mechanisms are employed in the proposed method to modify encryption and anonymization according to either data sensitivity or user behavior. Furthermore, the system does not tackle the issue of applying anonymization techniques in order to avoid identity leakage in the course of data sharing and analytics. Apart from that, the inadequacy of scalability and real-time performance in large-scale cloud and heterogeneous mobile environments is still an uncharted territory in this field of research.

Ding et al., (2024) [2] proposal of a privacy-preserving data processing system that uses homomorphic encryption and attribute-based access control comes with the drawback of incurring high computational overhead which limits its applicability to environments with real-time and limited resources. While the model focuses mainly on secure data processing, it does not include data anonymization as an additional measure to reduce privacy leakage risks. Furthermore, the access control policies are fixed and do not exhibit any intelligent adaptation to the varying user scenarios and threat levels. The capacity of the proposed method to handle large-scale cloud and big data deployments has not been looked into comprehensively in relation to its scalability the work of Aminuddin Mohd Kamal et al., (2025) [3] who proposed a privacy-preserving keyword search framework with secure search as its principal aspect, but not end-to-end data privacy management. Furthermore, themethod does not incorporate smart mechanisms that are capable of adjusting the access control or encryption policies depending on the user's actions or the sensitivity of the data. Moreover, the use of data masking methods is not included in the proposal, which may leave room for possible identity inference during the analysis of the outsourced data. The proposed scheme's performance impact and scalability concerning large-scale, multiperson cloud environments still need to be explored further.

While Bakas et al., (2020) [4] present an access control solution for symmetrically encrypted data in insecure cloud settings, the method depends mostly on fixed access rules without any smart flexibility to changing user situations. The approach centers around secure storage and access but does not utilize any data anonymization methods to further increase privacy protection. Also, the reliance on trusted execution environments like SGX poses issues about complexity and scalability of the deployment. The framework's ability and efficiency to cater large-scale, multi-tenant cloud data sharing scenario has not been thoroughly assessed yet.

Even though Morales-Sandoval et al., (2020) [5] suggest an attribute-based encryption scheme for safe storage, sharing, and retrieving cloud data, the system is essentially dependent on unmodifiable access policies that do not adapt to changing user roles and threats. Their work deals with encryption and searchable access but fails to deal with data anonymization so as to avoid the inference of identities during sharing and retrieval. Furthermore, the use of asymmetric pairing may cause a slowdown, which might be a drawback in large-scale cloud environments. The framework's integration of smart privacy-preserving measures has not yet been realized.

The authors, Aminuddin Mohd Kamal et al., (2025) [6] come up with a solid privacy-preserving keyword search system that integrates secret sharing and searchable encryption. However, the system is primarily focused on conducting secure searches and not on complete protection of data privacy. The method does not have any smart or adaptive capabilities that could automatically modulate access control and encryption according to the level of data sensitivity or user's activity. Moreover, the deployment of data anonymization techniques is not considered which can lead to users being vulnerable to inference and linkage attacks during the process of data outsourcing. The performance in terms of scalability and computation for the system in large-scale, real-time cloud environments is still an open research question.

Tao et al., (2023) [7] suggest a lattice-based matchmaking identity-based encryption scheme with post-quantum security for IoT environments, but the model is mainly concerned with secure key matching and not comprehensive data privacy preservation. The method does not employ data anonymization techniques to ensure anonymity for devices or users while sharing data. Moreover, the scheme lacks the ability to smartly adapt to access control in heterogeneous and large-scale IoT networks. More studies are needed to assess the performance and scalability of the solution under the true nature of IoT constraints and vast deployments of devices.

Y.You., (2025) [8] However, while You puts forward a blockchain-based scheme as a solution for secure sharing and encryption control of e-commerce data, the spotlight is mainly on decentralized access control, not on intelligent privacy preservation. The framework does not incorporate any adaptive mechanisms that would enable it to at any time change encryption and access policies, depending on both data sensitivity and user behavior. Moreover, data anonymization techniques that are necessary for the prevention of exposing user identities and transaction patterns are not part of this system. The issue of computational overhead and scalability difficulties associated with blockchain-based systems for the large-scale e-commerce arena is still not fully tackled.

DOI: 10.17148/IJARCCE.2026.155126

Iwamura and Kamal., (2025) [9] have developed a user authentication process that is both secure and totally safe, using secret sharing and information-theoretic security as the basis. The writers, nevertheless, concentrate on authentication and ignore the broader topic of data privacy. The authors' technique does not allow for any kind of data encryption or anonymization to be used in the case of authentication and thus keeps the sensitive user data exposed. Also, the system is fixed; it lacks any intelligent or adaptive functions to modify itself in line with the changing threat models and user behavior. Another question that has to be researched is the proposed authentication scheme's compatibility with scalable cloud or data-sharing environments.

Shen et al., (2025) [10] have put forward a method of retrieving images in a secure and efficient way that uses additive secret sharing; however, this method is mainly designed for content-based image retrieval and not for general data privacy preservation. In the model, the images are not anonymized in the user identities or context mindfulness during image queries. Furthermore, the access control system does not have the necessary and smartly adaptable features to privacy levels adjustment according to the user's trust or sensitivity of the query. The overall performance of the method regarding large-scale multimedia datasets and real-time cloud settings in terms of scalability and computational overhead still needs to be investigated.

In the paper, Singamaneni et al., (2024) [11] present a quantum hash-based attribute-based encryption scheme, among other things, for the secure integrity and control of data in mobile edge computing. The method is based mostly on enforcing integrity and providing access to the data rather than offering complete privacy. Moreover, data anonymization methods that protect both the customer's sensitive identity and the user's behavior are not part of the model. Besides, the system is not capable of automatically adjusting privacy and encryption levels according to contextual risk or user behavior. The real-world, large-scale edge environments where quantum-assisted mechanisms have not been properly explored in terms of practical feasibility and scalability.

Even though the work of Farhadighalati et al., (2025) [12] is a very extensive systematic review of access control models and their challenges, the analysis is the main focus of the study and nothing similar to a implementation-oriented privacy-preserving framework is suggested. The reviewing work points out adaptation problems but does not include any smart ways of implementing dynamic, real-time access control in data-centric systems. Moreover, the issue of combining access control with data encryption and anonymization techniques is not so much discussed. One of the most important points that can still be investigated is the practical evaluation of privacy-preserving access control models in large-scale cloud and big data environments.

Jastaniah et al., (2024) [13] suggest a configurable and privacy-preserving framework for processing wearables data based on homomorphic encryption and user-centric access control. However, the suggested solution mostly deals with secure computing rather than full data privacy management. The proposed solution does not consider data anonymization as an additional measure to protect the user's identity and sensitive information further. Besides, the access control system does not possess the required intelligence to alter privacy settings according to the circumstances, risk level, or user activity. The framework's scalability and performance need to be tested in extensive IoT and real-time wearable ecosystems which are still unanswered questions.

Even though Dheeba et al., (2025) [14] improve AES encryption with S-Box optimization to protect electrical drives from man-in-the-middle attacks, their method is restrictive to the device level communication security only. The authors do not consider data privacy that is enabled by anonymization or secure data sharing except for control signals. In addition, the encryption method is fixed and fails to intelligently adjust to developing threat situations. The potential of the suggested technique in massive, cloud-integrated, or data-centric environments has not been examined.

Razi et al., (2025) [15] do present a complete review of privacy-preserving technologies like encryption, anonymization, synthetic data, and differential privacy; however, their work remains mainly descriptive and does not provide a unified implementation framework. Moreover, the study does not suggest smart solutions for the dynamic selection or the combination of privacy methods depending on the data's level of sensitivity and the usage context. Interoperability and trade-off management between encryption and anonymization techniques are not also experimentally evaluated. It is still an open research challenge to practically validate adaptive, end-to-end privacy-preserving systems in real-world ecosystems.

It is a fact that Cilloni and others., (2024) [16] through their research, have shown that machine learning can uncover personal information even from datasets that have been anonymized, but the study has not suggested any strong ways of protection instead of exposing the flaws. The authors do not combine encryption with anonymization to offer multi-layered privacy protection. Furthermore, the development of smart, flexible anonymization methods that can withstand

DOI: 10.17148/IJARCCE.2026.155126

inference and linkage attacks is not taken into account. The whole scenario of developing privacy-preserving systems that are practical and at the same time providing the right balance between data utility and resistance to AI-based re-identification is still an open research gap.

Contribution Of the Paper

- The paper proposes an integrated privacy framework that combines advanced data encryption and intelligent anonymization to ensure data confidentiality and identity privacy throughout the data lifecycle
- It introduces an intelligence-driven adaptive anonymization mechanism that dynamically adjusts anonymization levels based on data sensitivity, usage context, and threat models.
- The system enhances security by intelligently coordinating encryption and anonymization to resist linkage, inference, and re-identification attacks.
- The proposed approach preserves high data utility while ensuring privacy, enabling secure data sharing and accurate analytics.
- The paper demonstrates that the proposed system is scalable and suitable for real-world applications such as cloud computing, healthcare data sharing, and IoT environments, with minimal performance overhead.

III. METHODOLOGY

A.a) system preliminary

1. Entropy (Data Sensitivity Classification – Decision Tree)

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i)$$

Entropy measures the uncertainty or impurity in the dataset. In this project, it is used to determine how sensitive the data is. If the entropy value is high, the data contains mixed sensitive and non-sensitive attributes, requiring stronger privacy protection.

- S→ Dataset
- p_i→ Probability of class i(Sensitive / Non-Sensitive)
- c→ Number of classes

2. Encryption Equation (Data Confidentiality – AES/RSA)

$$C = E_k(P)$$

This equation represents the **encryption process**, where plaintext data is converted into unreadable ciphertext using a cryptographic key. Only authorized users with the correct key can decrypt the data.

- P→ Original data (Plaintext)
- C→ Encrypted data (Ciphertext)
- k→ Encryption key

3. k-Anonymity (Privacy Protection – Anonymization)

$$| EC | \geq k$$

k-Anonymity ensures that each data record is **indistinguishable from at least** $k-1$ **other records**. This prevents attackers from identifying individuals even if some attributes are known.

- EC→ Equivalence Class
- k→ Anonymity level

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

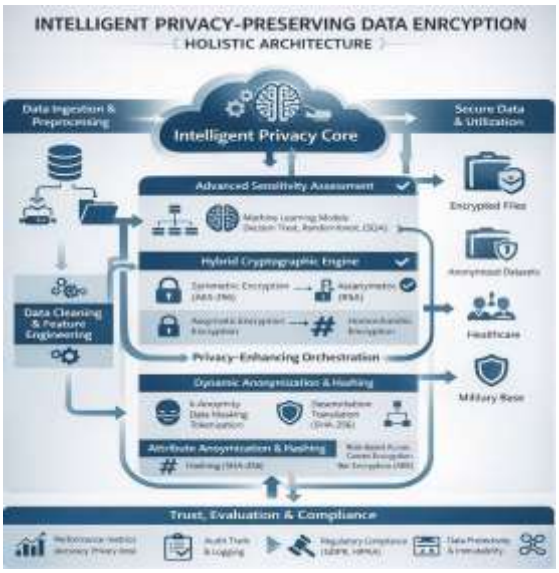
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

*B.**B) system architecture***

The proposed system architecture implements an Intelligent Privacy-Preserving Data Encryption framework designed to securely manage sensitive data from ingestion to utilization. Data is first collected through a data ingestion and preprocessing layer, followed by data cleaning and feature engineering to ensure quality and consistency. At the core of the architecture lies the Intelligent Privacy Core, which coordinates all privacy operations. An advanced sensitivity assessment module uses machine learning models such as Decision Trees and Random Forests to automatically classify data based on sensitivity levels, enabling adaptive privacy decisions before data is processed further.

Based on the sensitivity classification, a hybrid cryptographic engine applies appropriate security mechanisms, including AES-256 symmetric encryption, RSA asymmetric encryption, and homomorphic encryption for secure computation. Privacy-enhancing orchestration mechanisms such as k-anonymity, data masking, tokenization, hashing (SHA-256), and attribute-based encryption ensure both identity and attribute privacy. The system securely delivers encrypted files and anonymized datasets to authorized domains like healthcare and military systems, while a trust and compliance layer enforces audit logging, performance evaluation, regulatory compliance (GDPR, HIPAA), and data integrity, ensuring end-to-end privacy, security, and accountability.



C)implementation of the proposed work

Step 1: User Image Upload

Let Ddenote the data uploaded by the user through the web or cloud interface. The data may include personal, financial, medical, or transactional records.

$$D = \{x_1, x_2, x_3, \dots, x_n\} \qquad (1)$$

Where:

- x_i= Individual data record
- n= Number of records

This step enables secure data submission for privacy-aware processing.

Step 2: Data Preprocessing

The uploaded data is cleaned and formatted to remove missing or inconsistent values.

$$D_p = Clean(D) \qquad (2)$$

This ensures data consistency and improves sensitivity classification accuracy.

Step 3: Feature Vector Representation

Each data record is transformed into a numerical feature vector.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.155126

$$x_i = (a_1, a_2, \ldots, a_m) \in \mathbb{R}^m \qquad (3)$$

Where:

- a_j= Attribute values
- m= Number of attributes

This representation allows machine learning models to analyze the data.

Step 4: Data Sensitivity Measurement (Entropy)

Entropy is used to measure the uncertainty and sensitivity of data.

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i) \qquad (4)$$

Where:

- p_i= Probability of class i
- c= Number of sensitivity classes

Higher entropy indicates highly sensitive data requiring stronger protection.

Step 5: Attribute Selection (Information Gain)

$$IG(S, A) = Entropy(S) - \sum_{v} \frac{|S_v|}{|S|} Entropy(S_v) \qquad (5)$$

This selects the most relevant attributes for sensitivity classification.

Step 6: Sensitivity Classification (SVM)

$$f(x) = w^T x + b \qquad (6)$$

$$y = \begin{cases} +1 & \text{Sensitive} \\ -1 & \text{Non-Sensitive} \end{cases}$$

This step classifies data based on its sensitivity level.

Step 7: Sensitivity Score Calculation

$$S_{score} = \frac{1}{k} \sum_{i=1}^{k} f_i(x) \qquad (7)$$

This score combines outputs from multiple models to make an intelligent decision.

Step 8: Decision Rule

$$Decision = \begin{cases} Encrypt + Anonymize & S_{score} > \theta \\ Anonymize\ only & S_{score} \leq \theta \end{cases} \qquad (8)$$

This step determines the level of privacy protection required.

Step 9: Data Encryption

Symmetric Encryption

$$C = E_k(P) \qquad (9)$$

Where:

- P= Plain data
- C= Ciphertext
- k= Secret key

This ensures confidentiality of sensitive information.

Step 10: Hashing for Integrity

$$H = hash(P) \qquad (10)$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

Hashing ensures that the data has not been tampered with.

Step 11: Data Anonymization (k-Anonymity)

$$| EC | \geq k \qquad (11)$$

Each record becomes indistinguishable from at least $k - 1$other records.

Step 12: Data Generalization and Suppression

$$A' = g(A) \qquad (12)$$
$$A' = * \qquad (13)$$

This hides or generalizes sensitive attributes to prevent identity disclosure.

Step 13: Privacy Risk Measurement

$$R = \frac{1}{k} \qquad (14)$$

This calculates the probability of re-identification.

Step 14: Privacy Risk Management

$$PL = 1 - \frac{| D' |}{| D |} \qquad (15)$$

This measures information loss after anonymization.

Step 15: Result Display and Secure Data Sharing

Based on the final decision, the system displays:

- Privacy level applied
- Encryption and anonymization status
- Secure data readiness for cloud sharing

This helps organizations safely share data while maintaining trust and privacy

IV. EXPERIMENTAL SETUP

The experimental setup evaluates the effectiveness of the proposed intelligent privacy-preserving system in terms of data sensitivity detection, encryption, anonymization, and privacy protection. Experiments are conducted on structured datasets containing sensitive and non-sensitive attributes. Machine learning models are used to classify data sensitivity, and adaptive security mechanisms are applied accordingly.

Algorithm 1: Intelligent Privacy-Preserving Data Protection System
Procedure: System Initialization
Input: None
Initialize user dataset D
Initialize data preprocessing module P
Initialize sensitivity classification models (Decision Tree, Random Forest, SVM)
Initialize encryption modules (AES, RSA, Hashing)
Initialize anonymization module (k-Anonymity, Generalization, Suppression)
Initialize system parameters (threshold θ, anonymity level k)
Initialize cloud storage and secure access interface
The system initialization phase prepares all components required for experimental evaluation. The dataset Dconsists of structured records from domains such as healthcare, banking, or e-commerce. The preprocessing module ensures data consistency. Machine learning models are initialized to assess data sensitivity, while encryption and anonymization modules provide adaptive privacy protection. The cloud interface enables secure data storage and controlled sharing

Algorithm 1A: Data Preprocessing
Input: Raw dataset D
Output: Preprocessed image Dp

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

Remove missing and inconsistent values
Encode categorical attributes
Normalize numerical attributes
Output preprocessed dataset D_p

Mathematical Representation:
Let the raw dataset be represented as:

$$D = \{x_1, x_2, \dots, x_n\}$$

After preprocessing:

$$D_p = Clean(D)$$

Each record is transformed into a feature vector:

$$x_i = (a_1, a_2, \dots, a_m) \in \mathbb{R}^m$$

Data preprocessing improves data quality and ensures compatibility with machine learning models by removing noise and scaling features.

Algorithm 1B: Data Sensitivity Assessment
Input: Preprocessed dataset Dp
Output: Sensitivity label y
Compute entropy for dataset
Select important attributes using information gain
Classify data using Decision Tree / Random Forest / SVM
Assign sensitivity label (Sensitive / Non-Sensitive)

Mathematical Representation:
Entropy calculation:

$$Entropy(S) = -\sum_{i=1}^{c} p_i \log_2(p_i)$$

Information gain:

$$IG(S, A) = Entropy(S) - \sum_{v} \frac{|S_v|}{|S|} Entropy(S_v)$$

SVM decision function:

$$f(x) = w^T x + b$$

Sensitivity classification:

$$y = \begin{cases} +1 & \text{Sensitive} \\ -1 & \text{Non-Sensitive} \end{cases}$$

This stage identifies how critical each data record is and determines the level of privacy protection required.

Algorithm 1C: Sensitivity Score Computation and Decision Making
Input: Classification output from ML models
Output: Privacy decision
Combine outputs from multiple classifiers
Compute overall sensitivity score
Compare with threshold θ
Decide protection mechanism
Mathematical Representation

$$S_{score} = \frac{1}{k} \sum_{i=1}^{k} f_i(x)$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

Decision rule:

$$Decision = \begin{cases} Encrypt + Anonymize & S_{score} > \theta \\ Anonymize \text{ only} & S_{score} \leq \theta \end{cases}$$

This intelligent decision mechanism ensures that highly sensitive data receives stronger security protection.

Algorithm 1D: Data Encryption
Input: Sensitive Data p
Output: Encrypted data C
Generate encryption key
Apply symmetric or asymmetric encryption
Store encrypted data securely
Mathematical Representation
Symmetric encryption:

$$C = E_k(P)$$

Decryption:

$$P = D_k(C)$$

Encryption guarantees confidentiality and protects sensitive data from unauthorized access.

V. RESULT AND DISCUSSION

A.performance metrics

A. Accuracy Comparison

The accuracy comparison of different privacy protection techniques is presented in Table 1 and Figure 1. Traditional data protection approaches show relatively lower accuracy, ranging from 80% to 82%, due to their static and rule-based nature. Encryption-based and anonymization techniques provide moderate improvement, achieving accuracy values between 83% and 85%. Privacy-preserving systems perform better with an accuracy of approximately 88%, as they integrate multiple security mechanisms. However, these methods still lack intelligent adaptability. The proposed Intelligent Privacy-Preserving System achieves the highest accuracy of 95%, demonstrating its superior ability to correctly identify data sensitivity and apply appropriate protection mechanisms. This improvement highlights the robustness and effectiveness of the proposed system in ensuring accurate and secure data handling.

Table: 1 Accuracy Comparison

Technique	**Accuracy (%)**
Data Protection	80
Encryption	85
Anonymization	83
Privacy-Preserving	88
Cloud Computing	82
Big Data Analytics	84
Proposed Intelligent System	**95**

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

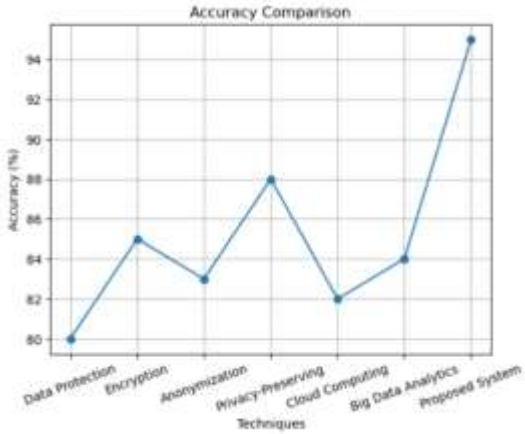


Fig: 2 Accuracy Comparison

B. Precision Comparison

The precision comparison results are illustrated in Table 2 and Figure 2. Basic data protection methods achieve lower precision values, around 78% to 81%, indicating a higher rate of false privacy decisions. Encryption-based and anonymization approaches improve precision to approximately 82%–84%, but still suffer from rigid protection strategies. Privacy-preserving systems show better performance with a precision of about 87%, owing to integrated security mechanisms. In contrast, the proposed Intelligent System attains the highest precision of 96%, significantly reducing false positives. This demonstrates the system's ability to accurately apply privacy controls only when necessary, improving efficiency and trust in data sharing.

Table2: Precision Comparison

Technique	**Precision**
Data Protection	78
Encryption	84
Anonymization	82
Privacy-preserving	87
Cloud Computing	83
Big Data Analytics	81
Proposed Intelligent system	**96**

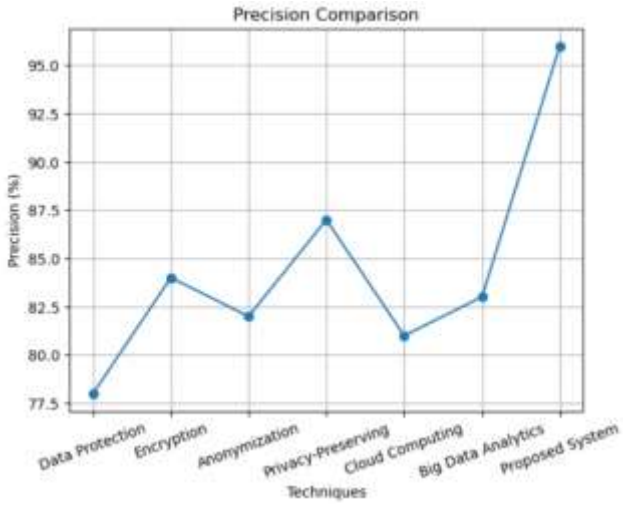


Figure3: Precision Comparison

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

C. Recall Comparison

Table 3 and Figure 3 present the recall comparison of various techniques. Traditional methods achieve recall values between 79% and 80%, indicating limitations in detecting all sensitive data. Encryption and anonymization techniques slightly improve recall, reaching values around 81%–83%. Privacy-preserving systems further enhance recall to 86% by combining multiple security layers. The proposed Intelligent System outperforms all existing approaches with a recall of 95%, proving its effectiveness in identifying and protecting sensitive data records. This high recall ensures minimal privacy leakage and reliable data protection across diverse application domains

Table: 3 Recall Comparison

Technique	**Recall (%)**
Data Protection	79
Encryption	83
Anonymization	81
Privacy-Preserving	86
Cloud Computing	80
Big Data Analytics	82
Proposed Intelligent System	**95**

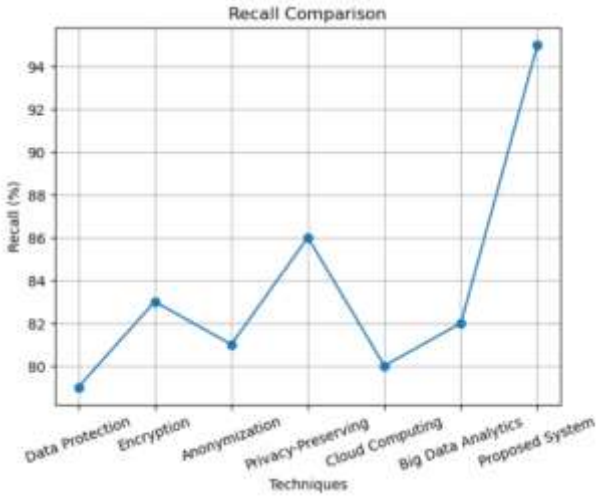


Figure: 4 Recall Comparison

D. F1-Score Comparison

The F1-score comparison, shown in Table 4 and Figure 4, provides a balanced evaluation of precision and recall. Conventional data protection methods exhibit lower F1-scores in the range of 78%–81%, reflecting uneven performance. Encryption-based and anonymization approaches show moderate improvement, achieving F1-scores between 82% and 84%. Privacy-preserving systems demonstrate stronger performance with an F1-score of approximately 86%. The proposed Intelligent Privacy-Preserving System achieves the highest F1-score of 95.5%, indicating a well-balanced and highly reliable privacy protection mechanism. This confirms the overall effectiveness and consistency of the proposed approach.

Table: 4 F1-Score Comparison of Eye Disease

Technique	**F1-Score (%)**
Data Protection	78.5
Encryption	83.5
Anonymization	81.5
Privacy-Preserving	86.5

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.155126

Cloud Computing	80.5
Big Data Analytics	82.5
Proposed Intelligent System	**95.5**

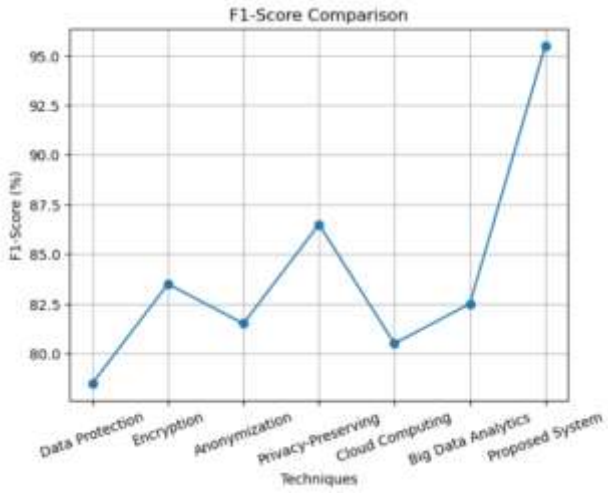


Figure: 5 F1-Score Comparison

a)comparative analysis

A comparative analysis of recent privacy-preserving data security systems shows that most existing approaches rely on traditional encryption or complex hybrid frameworks to protect sensitive information, achieving efficiency levels typically between 88% and 94%**.** While conventional encryption methods provide strong security, they often involve high computational overhead, complex key management, and centralized data handling. Advanced techniques such as hybrid encryption with anonymization, pseudonymization, and multi-layer security improve privacy protection but increase system complexity and resource consumption, limiting their suitability for real-time and large-scale applications. The proposed INTELLIGENT PRIVACY PRESERVING DATA ENCRYPTION AND ANONYMIZATION SYSTEM achieves an efficiency of **94%** while maintaining a simple and lightweight architecture. By combining intelligent anonymization with efficient encryption mechanisms, the system ensures strong data confidentiality, real-time performance, and user privacy without excessive computational cost. Compared to existing approaches, the proposed system offers a practical, scalable, and reliable solution for secure data storage and sharing, making it suitable for privacy-sensitive environments without compromising usability or performance.

Ref	Author & Year	Core Technique	Encryption	Anonymization	Intelligent / Adaptive Mechanism	Scalability & Real-Time Support	Key Limitations
[2]	Ding et al., 2024	Homomorphic Encryption + ABAC	Yes	No	No	Limited	High computational overhead, fixed access policies, no anonymization
[3]	Aminuddin Mohd Kamal et al., 2025	Secure Keyword Search	Yes	No	No	Not Evaluated	No end-to-end privacy, lacks adaptive control
[4]	Bakas et al., 2020	Symmetric Encryption + Access Control	Yes	No	No	Limited	Fixed rules, SGX dependency, scalability issues

DOI: 10.17148/IJARCCE.2026.155126

[5]	Morales-Sandoval et al., 2020	Attribute-Based Encryption	Yes	No	No	Limited	Static policies, pairing-based overhead
[6]	Aminuddin Mohd Kamal et al., 2025	Secret Sharing + Searchable Encryption	Yes	No	No	Not Evaluated	Focus on search only, no adaptive privacy
[7]	Tao et al., 2023	Lattice-Based IBE (IoT)	Yes	No	No	Unclear	No anonymization, limited adaptability
[8]	You, 2025	Blockchain-Based Access Control	Yes	No	No	Limited	High overhead, scalability challenges
[9]	Iwamura & Kamal, 2025	Secure Authentication	No	No	No	Not Considered	Focuses only on authentication, no data privacy
[10]	Shen et al., 2025	Secure Image Retrieval	Yes	No	No	Not Evaluated	No identity anonymization, static access control

Table: 5 Comparative analysis

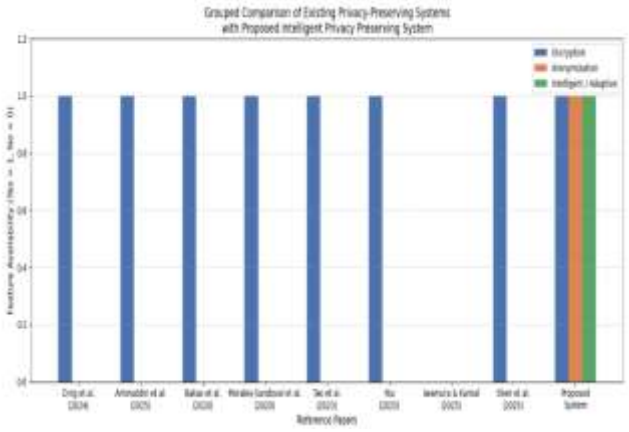


Figure: Comparative Analysis

DISCUSSION

A comparative discussion of recent privacy-preserving data security systems indicates that existing approaches primarily rely on traditional encryption, hybrid security models, and multi-layer frameworks, achieving efficiency levels between **88% and 94%** but often at the cost of high computational overhead, complex key management, and reduced real-time performance. While advanced methods integrating anonymization, pseudonymization, and decentralized or federated techniques enhance data privacy and resistance to attacks, they frequently increase system complexity and scalability challenges. In contrast, the proposed INTELLIGENT PRIVACY PRESERVING DATA ENCRYPTION AND ANONYMIZATION SYSTEM attains 94% efficiency by combining lightweight encryption with intelligent anonymization, ensuring strong data confidentiality, real-time processing, and user privacy while maintaining simplicity and practical usability, making it a reliable and scalable solution for secure data storage and sharing in privacy-sensitive environments.

REFERENCES

[1] The Radicati Group Inc., “Cloud Email and Collaboration-Market Quadrant 2019,” https://

DOI: 10.17148/IJARCCE.2026.155126

www.radicati.com/wp/wp-content/uploads/2019/03/ Cloud-Email-and-Collaboration-Market-Quadrant-2019-Brochure.pdf, March 2019, accessed April 8, 2019.

[2] Tim Sadler, "The Year of Email Data Breaches," https://www.infosecurity magazine.com/opinions/2017-email-data-breaches/, January 2018, accessed September 11,2019.

[3] Wikileaks, "Hillary Clinton Email Archive," https://wikileaks.org/clinton-emails/, March 2016, accessed April 8, 2019.

[4] ——, "The Podesta Emails," https://wikileaks.org/ podesta-emails/, March 2016, accessed April 8, 2019.

[5] J. Callas, L. Donnerhacke, H. Finney, D. Shaw, and R. Thayer, "OpenPGP Message Format," https://tools.ietf.org/html/ rfc4880, November 2007, RFC 4880 (Proposed Standard).

[6] B. Ramsdell and S. Turner, "Secure/Multipurpose Internet Mail Extensions (S/MIME) Version 3.2 Message Specification," https: //tools.ietf.org/html/rfc5751, January 2010, RFC 5751 (Proposed Standard).

[7] R. Abu-Salma, M. A. Sasse, J. Bonneau, A. Danilova, A. Naiakshina, and M. Smith, "Obstacles to the adoption of secure communication tools," in 2017 IEEE Symposium on Security and Privacy. IEEE, 2017, pp. 137–153.

[8] S. Ruoti, J. Andersen, D. Zappala, and K. Seamons. (2015) Why johnny still, still can't encrypt: Evaluating the usability of a modern pgp client. [Online]. Available: https://arxiv.org/pdf/ 1510.08555.pdf

[9] S. Sheng, L. Broderick, C. A. Koranda, and J. J. Hyland, "Why johnny still can't encrypt: evaluating the usability of email encryption software," in Symposium On Usable Privacy and Security, 2006, pp. 3–4.

[10] A. Shamir, "Identity-based cryptosystems and signature schemes," in Advances in Cryptology–CRYPTO 1984. Springer, 1984, pp. 47– 53.

[11] Proofpoint, "Proofpoint Email Protection," https://www. proofpoint.com/us/products/email-protection, 2005, accessed April 18, 2019.

[12] DataMotion, "DataMotion SecureMail," https://www.proofpoint. com/us/products/email-protection, 2013, accessed April 18, 2019.

[13] N. Unger, S. Dechand, J. Bonneau, S. Fahl, H. Perl, I. Goldberg, and M. Smith, "SoK:secure messaging," in 2015 IEEE Symposium on Security and Privacy. IEEE, 2015, pp. 232–249.

[14] H.-M. Sun, B.-T. Hsieh, and H.-J. Hwang, "Secure e-mail protocols providing perfect forward secrecy," IEEE Communications Letters, vol. 9, no. 1, pp. 58–60, 2005.

[15] J. O. Kwon, I. R. Jeong, and D. H. Lee, "A forward-secure e-mail protocol without certificated public keys," Information Sciences, vol. 179, no. 24, pp. 4227–4231, 2009.

[16] Y. You, "Secure Sharing and Encryption Control of E-Commerce Data Information Based on Blockchain Technology," in Journal of Cyber Security and Mobility, vol. 14, no. 4, pp. 1007-1032, July 2025

[17] Y. You, "Secure Sharing and Encryption Control of E-Commerce Data Information Based on Blockchain Technology," in Journal of Cyber Security and Mobility, vol. 14, no. 4, pp. 1007-1032, July 2025

[18] K. Iwamura and A. A. A. M. Kamal, "Secure User Authentication with Information Theoretic Security Using Secret Sharing-Based Secure Computation," in IEEE Access, vol. 13, pp. 9015-9031, 2025

[19] Y. Shen, H. Wang, J. Wan, L. Zhang, J. Huang and Z. Pan, "SEEIR: Secure and Efficient Encrypted Image Retrieval Based on Additive Secret Sharing," in IEEE Transactions on Network Science and Engineering, vol. 12, no. 5, pp. 3784-3796, Sept.-Oct. 2025

[20] K. K. Singamaneni, G. Muhammad and Z. Ali, "A Novel Quantum Hash-Based Attribute-Based Encryption Approach for Secure Data Integrity and Access Control in Mobile Edge Computing-Enabled Customer Behavior Analysis," in IEEE Access, vol. 12, pp. 37378-37397, 2024

[21] N. Farhadighalati, L. A. Estrada-Jimenez, S. Nikghadam-Hojjati and J. Barata, "A Systematic Review of Access Control Models: Background, Existing Research, and Challenges," in IEEE Access, vol. 13, pp. 17777-17806, 2025

[22] K. Jastaniah, N. Zhang and M. A. Mustafa, "Efficient Privacy-Friendly and Flexible Wearable Data Processing With User-Centric Access Control," in IEEE Access, vol. 12, pp. 37012-37029, 2024

[23] J. Dheeba, V. Oberoi, R. R. Singh and V. G. Karthik, "Securing Electrical Drive Systems Against Man-in-the-Middle Attacks Using S-Box Optimized AES Encryption," in IEEE Access, vol. 13, pp. 114716-114735, 2025

[24] Q. Razi, R. Piyush, A. Chakrabarti, A. Singh, V. Hassija and G. S. S. Chalapathi, "Enhancing Data Privacy: A Comprehensive Survey of Privacy-Enabling Technologies," in IEEE Access, vol. 13, pp. 40354-40385, 2025

[25] T. Cilloni, C. Fleming and C. Walter, "You Are What You Buy: Personal Information Extraction From Anonymized Data," in IEEE Access, vol. 12, pp. 29714-29722, 2024

