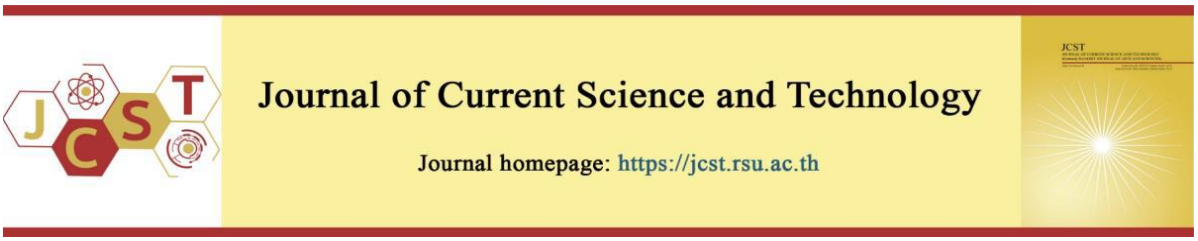
Journal of Current Science and Technology, September-December 2024 Vol. 14 No. 3, Article 57

ISSN 2630-0656 (Online)

Cite this article: Shobha, P., & Nalini, N. (2024). Genomic data fusion using paillier cryptosystem. *Journal of Current Science and Technology*, *14*(3), Article 57. https://doi.org/10.59796/jcst.V14N3.2024.57



Genomic Data Fusion using Paillier Cryptosystem

Shobha P.*, and Nalini N.

Computer Science and Engineering Department, Nitte Meenakshi Institute of Technology (Affiliated to Visvesvaraya Technological University), Bengaluru, Karnataka, 560064, India

*Corresponding author: E-mail; shobha.kp@gmail.com

Received date 19 November 2023; Revised 12 March 2024; Accepted 28 April 2024
Published online 1 September 20024

Abstract

The proposed work performs secure data fusion using homomorphic encryption, specifically the Paillier cryptosystem. The Paillier cryptosystem allows computation to be performed on encrypted data without decrypting it first, thus ensuring the privacy and security of the computation. The experiment measures the algorithm's performance based on execution time, memory usage, security, accuracy, and scalability. The data-level Paillier cryptosystem approach is generally slower than the feature-level fusion method due to its more complex operations and computations. Scalability is limited by the time required for encryption, homomorphic addition, and decryption. Improving scalability can be achieved by parallelizing the encryption and decryption steps, optimizing the homomorphic addition algorithm, or using more efficient cryptographic primitives. The article compares the performance of the Paillier cryptosystem with differential privacy in terms of their advantages and disadvantages. By adopting a preemptive approach to data fusion security, healthcare organizations can minimize the risk of data breaches and protect patient privacy. Data fusion security is an important factor when dealing with medical records. In the field of medical records, data fusion refers to the method of combining multiple sources of data into a distinct record. This can include data from electronic health records (EHRs), medical imaging devices, wearable devices, and other sources. There are several security considerations that must be addressed when fusing data from multiple sources.

Keywords**:** *artificial intelligence; data Fusion; differential privacy; data privacy; homomorphic; medical; paillier cryptosystem; radar data set*

Symbols

Symbol	Description	Symbol	Description
S_1	First data source	L	Large prime
S_2	First data source	M	Large prime
S_1^I	Perturbed version of S_1	N	Product of L and M
$S_1 d$	Vector representation of S_1	gen	Generator
$C(\beta)$	Perturbation term	$randm$	Random number
Δ	Parameter defining the level of perturbation	CYP	Ciphertext
y	Random variable following a normal distribution	pc_fuse	Fused data
P	Protection mechanism	$feature$	Decrypted feature
λ	Result of least common multiple operation on two large primes	μ	Result of mathematical operation involving a generator

1. Introduction

Artificial Intelligence (AI) systems often process large volumes of data from multiple sources, such as sensors, databases, and external Application Programming Interface (external APIs), to gain insights and make predictions. However, the accuracy and usefulness of these predictions can be limited by incomplete or conflicting data. Data fusion techniques aim to overcome these limitations by combining data from numerous sources to create a completer and more accurate picture of the situation at hand.

There are various techniques used in data fusion, including sensor fusion. This engages combining data from multiple sensors to obtain more accurate measurements of a particular phenomenon. Feature fusion involves combining features obtained from multiple data sources to create a more comprehensive representation of the data (Singh, & Barde, (2024). Decision fusion involves combining multiple decisions made by different AI models to make a final decision or prediction (Dai et al., 2023; Alipour et al., 2023). Data fusion using machine learning trains a machine learning model to integrate data from multiple sources and make predictions based on the combined data. Data fusion is an important aspect of AI because it allows machines to make more accurate and informed decisions by leveraging data from multiple sources. Data fusion is used in medical imaging to improve the accuracy of diagnoses. By combining different imaging modalities, such as Magnetic resonance imaging (MRI), Computed tomography (CT), and Positron emission tomography (PET), doctors can obtain a complete picture of the patient's condition. This can help them make a more accurate diagnosis and plan the appropriate treatment. Data fusion can be used to combine information from different electronic health records (EHRs) to create a more complete patient profile. This can lead to better diagnoses and more effective treatments.

Wearable devices, such as fitness trackers and smartwatches, can collect a wide range of data on the wearer's health, such as heart rate, sleep patterns, and activity levels. By combining this data with other health information, such as EHRs, doctors can gain a more complete understanding of the patient's health and identify potential health risks. Data fusion holds immense potential to revolutionize healthcare by providing doctors with more complete and accurate information on patients, a critical caveat prevails. It is imperative to ensure that the handling of patient data adheres to stringent standards of security and ethics, safeguarding patient privacy. This ethical consideration underscores the responsible use of data, aligning with regulatory frameworks to ensure the integrity and confidentiality of sensitive healthcare information.

Data fusion can also be used in genomics to combine information from different genetic tests. By integrating genetic data with other health information, doctors can gain a better understanding of the patient's risk for certain diseases and tailor treatments to their specific genetic makeup.

Data fusion has the potential to revolutionize healthcare by providing doctors with more complete and accurate information on patients. However, it is important to ensure that patient data is handled in a secure and ethical manner to protect patient privacy.

2. Related Work

Data fusion is the process of integrating multiple sources of data to provide a comprehensive and accurate understanding of a particular situation. In the medical field, data fusion can be used to combine data from different sources, such as EHRs, medical imaging systems, and medical devices, to gain a more complete picture of a patient's health status (Steyaert et al., 2023; Albahri et al., 2023; Anita, & Kumaran, 2023). Fusion techniques are used for medical diagnosis. The presence of nonlinear deterministic structures in brain electrical activity, and their dependence on recording region and brain state (Rainio et al., 2023) are studied. Data fusion approach is applied for accurate classification of electro-cardiogram signals. Data mining techniques are used for medical image classification. While data fusion can be useful in improving medical care, it can also raise security concerns. Medical data is highly sensitive and valuable, making it a prime target for cybercriminals. There are several strategies that can be employed to ensure data security. One approach is to use encryption and access controls to secure the data and limit access to authorized personnel. Another approach is to implement robust authentication and identity management systems to prevent unauthorized access to the data. It is important to have a comprehensive security plan that includes regular security audits and vulnerability assessments, as well as contingency plans in case of a security breach. This involves backing up data, establishing disaster recovery plans, and having protocols in place to detect and respond to security incidents. The articles demonstrate that data fusion of wearable sensor

data is a promising approach for human activity recognition, with the potential to improve accuracy and enable new applications in various fields (Albahri et al., 2023; Jemili, 2023; Mohsen et al., 2022). The proposed approaches vary in terms of the sensors used, the fusion techniques applied, and the machine learning algorithms employed, highlighting the diversity of approaches in this area of research. A novel deep learning-based method for identifying disease-related genes by integrating multiple types of genomic data (Al-Hawawreh, & Hossain, 2023; Nguyen et al., 2023). The method combines a deep neural network with a Bayesian network to model the dependencies between genes and different types of genomic data. The authors evaluate the proposed method on several real-world datasets and show that it outperforms other state-of-the-art methods in terms of identifying disease-related genes.

The frameworks and methods for multimodal medical image fusion using deep learning techniques (Mergin, & Premi, 2023; Babu, & Narayana, 2023; Sunitha et al., 2022) are also examined. They propose different architectures and approaches for combining data from different medical imaging modalities, such as CT and MRI, to improve the accuracy and comprehensiveness of medical diagnosis. These methods involve the use of convolutional neural networks (CNNs), sparse representation, fuzzy entropy, and multiscale analysis to learn and extract features from the input images and to fuse them into a single output image. The proposed methods are evaluated and compared based on various metrics, including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and fusion quality index (FQI), to assess their effectiveness in improving image quality and clinical performance.

The current state of research on data fusion techniques for EHR analysis implies future directions for research in this area (Raman et al., 2023; Liang et al., 2022). Data fusion in the medical field involves techniques such as deep learning-based frameworks, multi-sensor data fusion, and machine learning techniques. These techniques address various challenges related to medical data integration, such as accuracy, reliability, and efficiency, and propose novel solutions to improve medical diagnosis, monitoring, and treatment.

.

3. Proposed Method

Genomic data fusion is the integration of multiple genomic data types, such as DNA sequencing, gene expression, and epigenetic data to provide a more comprehensive assessment of the underlying biological mechanisms. Despite the potential benefits of genomic data fusion, there are various issues that need to be addressed. Genomic data fusion often involves integrating data from different sources with varying levels of quality and resolution, which makes it difficult to harmonize the data for analysis. To address this issue, standardization of data formats and metadata could help improve data integration and interoperability.

As the amount of genomic data being generated continues to grow exponentially, scalable and efficient data fusion methods are needed. This requires the development of new computational algorithms and tools that can handle large-scale data sets and complex data structures. Integrating multiple data types can result in complex and heterogeneous data sets, making it challenging to interpret the results. More effective visualization and data exploration tools are needed to help researchers better understand the underlying biological processes. Some data types may not be available for all samples, leading to missing data.

This can affect the accuracy of the results and make it difficult to draw meaningful conclusions. Addressing missing data requires the development of new imputation methods accurately estimate missing values. Genomic data contains sensitive information, and there are apprehensions about data privacy and security. To address these concerns, there is a need for new privacy-preserving methods that can enable secure sharing and analysis of genomic data fusion will be critical for understanding the full potential of this approach and for advancing understanding of complex biological systems. Data fusion in the medical field requires a secure process to protect patient privacy and prevent unauthorized access to sensitive information. Here are some key steps to ensure secure data fusion in the medical field.

Steps to ensure that patient data is de-identified or anonymized before it is integrated. It is also important to establish clear policies and procedures for handling patient data and to limit access to patient data to authorized personnel only. During the transfer of data between different sources, it is important to use secure methods to ensure that the data is not intercepted or compromised. This may involve using encryption, firewalls, and other security measures to protect data in transit.

To prevent unauthorized access to patient data, it is important to implement access controls that limit who can view and manipulate the data. This

may involve using role-based access controls, authentication, and other security measures to ensure that only authorized personnel can access patient data. To ensure that patient data is being used appropriately and securely, it is important to monitor data access and use. This may involve logging data access and use, using analytics tools to identify unusual patterns of data access or use, and regularly auditing data access and use to ensure compliance with policies and procedures. To ensure data fusion in the medical field is secure and compliant with relevant regulations and standards, it is important to establish a data governance framework. This may involve developing policies and procedures for data handling, establishing a data stewardship program, and regularly reviewing and updating these policies and procedures to ensure that they remain current and effective. A secure data fusion block diagram typically consists of several components working together to ensure the secure aggregation of data from multiple sources.

The Paillier cryptosystem (PC) is applied to perform secure data fusion (Figure 1). Paillier cryptosystem, a probabilistic public-key encryption method, supports homomorphic addition and multiplication. It enables computations directly on ciphertexts, thereby maintaining data privacy. This feature allows operations on encrypted data without requiring access to plaintext.

3.1 Data Sources

These are the various sources of data that are being collected, such as sensors, databases, medical records, radar data set, or other systems. Each source may have different types of data, different data formats, or different levels of sensitivity. In the proposed work, the subcellular and RNA data are used.

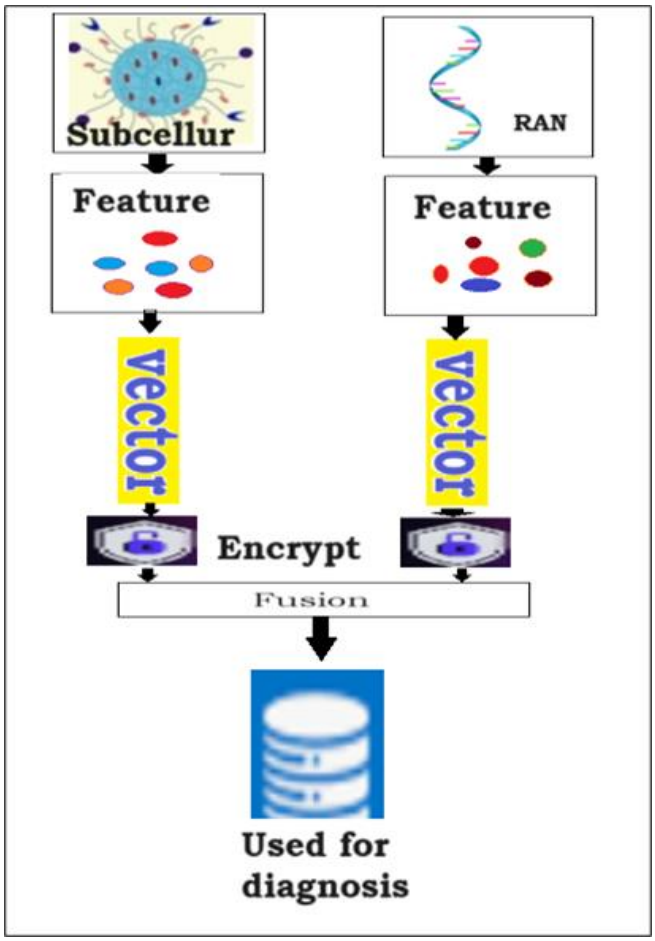


Figure 1 Framework for Secure Data Fusion using Homomorphic Encryption.

3.2 Data Pre-processing

This involves standardizing the data collected from various sources, which may include converting data into a common format or normalizing data for consistency. Pre-processing may also include basic data cleaning and filtering to remove any irrelevant or redundant information. Medical data is often complex and heterogeneous. Pre-processing steps may include data cleaning, anonymization, feature extraction, and normalization. The extracted features are represented in the form of vectors.

3.3 Encryption

Once the data is pre-processed, it is encrypted to protect it from unauthorized access or tampering. This can be done using standard encryption techniques, such as symmetric or asymmetric encryption, depending on the level of security required.

The Paillier cryptosystem is a public-key cryptosystem that can be used to securely fuse data from multiple sources (Ifzarne et al., 2023; Tsai et al., 2022). It is based on the homomorphic properties of the Paillier cryptosystem, which allow for the addition of encrypted data without decrypting it. This makes it ideal for data fusion applications, where data from multiple sources needs to be combined without revealing the individual sources. The Paillier cryptosystem is based on the concept of homomorphic encryption, which allows for the addition of encrypted data without decrypting it. Data from multiple sources can be combined without revealing the individual sources. The Paillier cryptosystem uses a public-key encryption scheme.

3.4 Data Fusion

The encrypted data is then combined or fused to form a unified data set that can be used for analysis or processing. This process may involve combining data from multiple sources into a single data set or analyzing the data to extract meaningful insights. This requires the use of encryption keys to unlock the data and ensure its integrity. The data is made available to authorized users or systems for analysis or use. Access control methods are put in place to restrict access to sensitive data and ensure that only authorized users can access it.

Secure data fusion involves the integration of data from multiple sources while ensuring that the data remains secure and private. This can be achieved through various cryptographic techniques, such as homomorphic encryption, secure multi-party computation, and differential privacy by applying equations. (*i* - *iv*).

3.5 Decryption

The fused data is then decrypted to reveal the original data from each source.

To implement the secure fusion let us assume two data sources, $S_1$and S_2, given a protection P which takes input s_1 and produces output $O(s_1)$, say P satisfies differential privacy, if for any pair S_1 and (S_1, $O(S_1)$,), and all possible subsets $S \in Range(O)$,

$$DP[P(S_1) \in S] \leq \exp(\varepsilon)\, DP\left[P\left(O(S_1)\right) \in S\right] \quad \text{(i)}$$

The source S_1 is represented by vector S_{1d}, The output is represented by.

$$S^{I}_{1d} = S_{1d} + C(\beta)\ \Delta/y,\ y \sim N\_d\ \ (0,P) \quad \text{(ii)}$$

$$\text{where sup} \parallel P\text{-}1/\,2\,(S_{1d} - S'_{1d}) \parallel 2 \leq \Delta \quad \text{(iii)}$$

In PC the Key generation involves randomly selecting two large primes L and M, where L|=|M|=|K|. Then calculating λ=LCM (L-1, M-1). Defined by a function L(V)=V-1/N, where N=L. M. Choose a generator gen $\in N_2$, and calculate μ = (geny mod N_2) - 1 mod N. The public key is (N, gen), and the corresponding private key is.

$$C(\beta) \geq \sqrt{2}\ \ \log\ [\![2 / \beta \quad \text{(iv)}$$

$$(\lambda, \mu) \quad \text{(v)}$$

To Encrypt the feature, f_1, where $f_1 \in ZN$, choose a random number $randm \in ZN^*$, then gcd(randm, N) =1.

$$CYP_1 = ENC(f_1)) = gen * randm\ N \bmod N_2 \quad \text{(v)},$$

The secured fused data is represented by.

$$pc_{fuse} = CYP_1 + CYP_2 \quad \text{(vii)}$$

Decryption. Given the ciphertext $C \in ZN$, the corresponding message is decrypted with the private key (λ, μ) as

$$\text{Feature} = DEC(CYP) = L\ (CYP\ \lambda \bmod N_2\ )\mu \bmod N \quad \text{(viii)}$$

Public and private key pairs are generated using (*v*). The public key can be used to encrypt (*vi*) to generate df the medical records, and the private key can be used to decrypt (*vii*) them. Once the keys are

generated, medical records can be encrypted using the public key. The encrypted records can then be combined using the homomorphic property of the Paillier cryptosystem. This property allows the encrypted records to be added together without decrypting them.

4. Results and Discussion

Homomorphic Encryption and the Paillier Cryptosystem enable computation on encrypted data without decryption. The Paillier cryptosystem facilitates homomorphic addition for secure data fusion, preserving privacy and security in scenarios where raw data sharing is restricted.

4.1 Performance Comparison with Differential Privacy

Experiments are carried out to perform secure data fusion using homomorphic encryption and the Paillier cryptosystem. Homomorphic encryption is a cryptographic technique that allows computation on encrypted data without decrypting it. The Paillier cryptosystem is used to perform homomorphic addition on encrypted data from two sources. This allows the data to be fused without revealing any individual data points to the other party.

The resulting sum can be decrypted only by the party who holds the private key, ensuring the privacy and security of the computation. This technique is useful in scenarios where data from multiple sources need to be combined, but privacy concerns or data ownership prevent the sharing of raw data. These are just a few examples, and there are many other types of data that can be used for data fusion. The choice of which data to use depends on the specific research question and the available data resources. The algorithm's performance is measured using execution time, memory usage, security, accuracy, and scalability.

Table 1 shows sample values resulting from fusion using Differential Privacy (Al-Hawawreh, & Hossain, 2023; Zhang et al., 2023a; Zhang et al., 2023b; Qi et al., 2023) and Paillier cryptosystem. The execution time of the data-level Paillier cryptosystem is higher than the execution time of the feature-level fusion method because of the Encryption and decryption of data: In the data-level fusion approach, the data needs to be encrypted and decrypted using the Paillier cryptosystem. Encryption and decryption are computationally intensive processes that can slow down the execution time of the program. In contrast, the feature-level fusion method does not require any encryption or decryption of data. The second reason is homomorphic addition. The data-level fusion approach uses homomorphic addition to combine the encrypted data. Homomorphic addition involves performing mathematical operations on encrypted data without decrypting them. Homomorphic operations are computationally intensive and can lead to higher execution times. In contrast, the feature-level fusion method only requires simple mathematical operations like addition and division. The third reason is the number of operations performed. In the data-level fusion approach, each data point needs to be encrypted, decrypted, and then combined using homomorphic addition. This involves many operations for each data point, which can increase execution time. In contrast, the feature-level fusion method requires only a few operations per data point. Therefore, the data-level Paillier cryptosystem approach is generally slower than the feature-level fusion method because it involves more complex operations and computations.

Table 1 Data Fusion

Source 1	Source 2	Differential Privacy	Paillier cryptosystem
1	6	6.437678	7
2	7	8.427249	9
3	8	12.455011	11
4	9	12.737888	13
5	10	15.880387	15

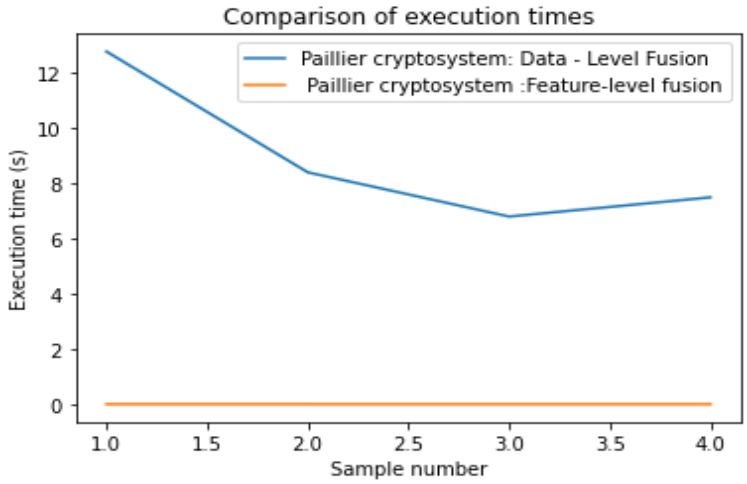


Figure 2 Execution time

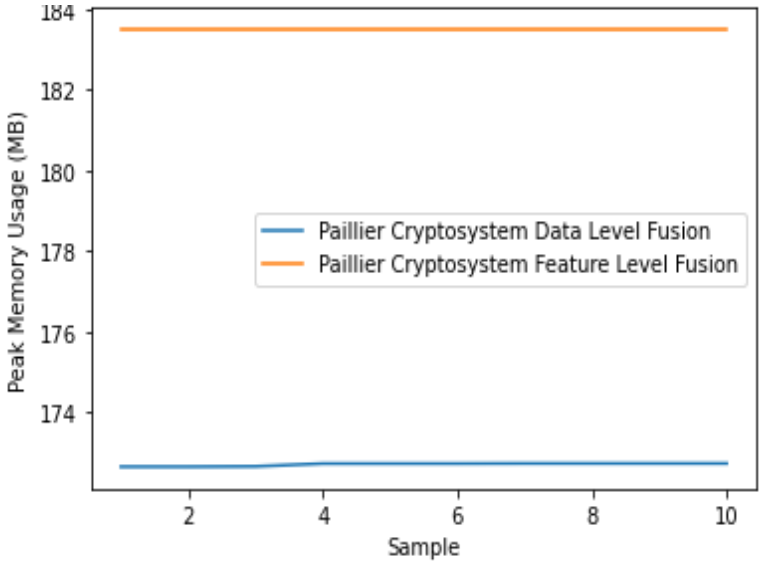


Figure 3 Memory usage of Paillier Cryptosystem Data Fusion**.**

The execution time (Figure 2) of the data-level Paillier cryptosystem is higher than that of the feature-level fusion method because in the data-level fusion approach, the data needs to be encrypted and decrypted using the Paillier cryptosystem. Encryption and decryption are computationally intensive processes that can slow down the program's execution time.

In contrast, the feature-level fusion method does not require the encryption or decryption of data. The data-level fusion approach uses homomorphic addition to combine the encrypted data. Homomorphic addition involves performing mathematical operations on encrypted data without decrypting it. Homomorphic operations are computationally intensive and can lead to higher execution times. In contrast, the feature-level fusion method requires only simple mathematical operations like addition and division.

In the data-level fusion approach, each data point needs to be encrypted, decrypted, and then combined using homomorphic addition. This involves many operations for each data point, which can increase the execution time. In contrast, the feature-level fusion method requires only a few operations per data point. Therefore, the data-level Paillier cryptosystem approach is generally slower than the feature-level fusion method because it involves more complex operations and computations. Scalability is limited by the time required for encryption,

homomorphic addition, and decryption. If the size of the data or the number of sources is small, the performance may be acceptable, but if the size is large, the execution time may become prohibitively long. To improve scalability, one could consider parallelizing the encryption and decryption steps, optimizing the homomorphic addition algorithm, or using more efficient cryptographic primitives. The Paillier cryptosystem is used for secure data aggregation or fusion, while differential privacy is used for preserving privacy in statistical analysis. The performance of the two techniques is compared in terms of their advantages and disadvantages. The Paillier cryptosystem is computationally intensive and requires more resources than differential privacy. It is suitable for situations where the data owner wants to share data with others while keeping it private. It offers strong security guarantees and enables data fusion without revealing the underlying data. On the other hand, differential privacy offers a more lightweight approach to preserving privacy in data analysis. It adds noise to the data to protect individual privacy while still allowing statistical analysis. Differential privacy is more suitable for situations where data needs to be analyzed in a centralized or distributed manner while preserving privacy. The choice of technique depends on the specific requirements of the use case, and a thorough analysis of the advantages and disadvantages of each technique should be performed before choosing one over the other.

The Paillier cryptosystem involves computationally intensive operations such as encryption, decryption, and homomorphic addition. These operations can take a significant amount of time to execute, especially when dealing with large datasets. The Paillier cryptosystem requires data to be encrypted and decrypted, which incurs additional overhead in terms of computation time and memory usage (Figure 3). The time taken (Figure 4) by the Paillier cryptosystem for data fusion may be higher than the time taken by differential privacy methods because of computationally intensive operations. The Paillier cryptosystem uses homomorphic addition to combine the encrypted data. Homomorphic operations can be computationally expensive, especially when dealing with large datasets. In contrast, differential privacy methods typically involve simpler operations, such as adding noise to the data or using randomized responses. These operations are less computationally intensive and can be performed more quickly. However, it is worth noting that the choice of data fusion method depends on several factors such as data sensitivity, the desired level of privacy, and the specific use case. Differential privacy methods may be more appropriate for some scenarios, while the Paillier cryptosystem may be more suitable for others.

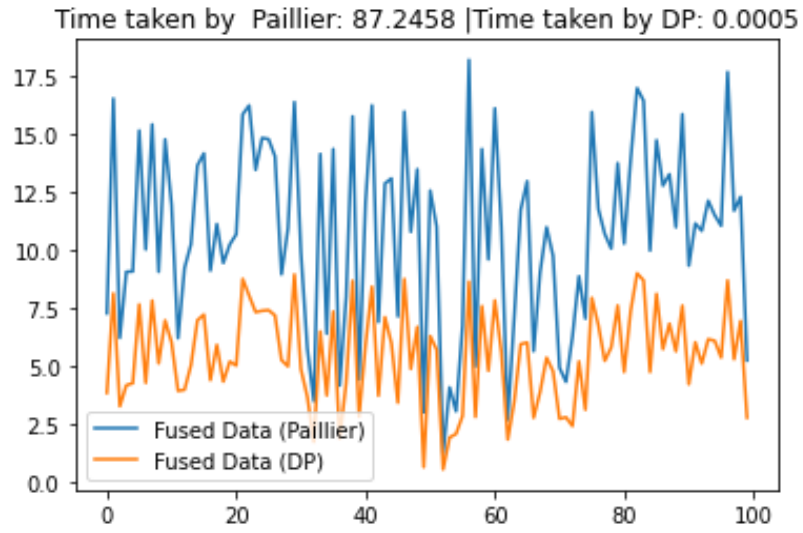


Figure 4 Comparison of time taken by Paillier vs Differential Privacy.

Table 2 Mean Square Error of Paillier and Differential Privacy.

Method	Mean Squared Error
Paillier	32.597
Differential Privacy	3.8964

The mean square error (MSE) is a metric used to measure the difference between the original data and the data after it has been processed. In the context of data fusion, MSE can be used to compare the accuracy of different data fusion methods. In the case of the Paillier cryptosystem and differential privacy (Table 2), both methods are used to perform secure data fusion. However, the Paillier cryptosystem uses homomorphic encryption to combine encrypted data, while differential privacy adds random noise to the data before combining it. Homomorphic encryption allows for secure computation on encrypted data without the need to decrypt it. However, homomorphic operations are computationally expensive and can result in a higher MSE. On the other hand, differential privacy adds random noise to the data, which can result in a lower MSE. The amount of noise added by differential privacy can be adjusted to balance the privacy and accuracy requirements. This makes differential privacy more flexible in terms of balancing privacy and accuracy compared to homomorphic encryption. Therefore, the MSE of the Paillier cryptosystem may be higher than that of differential privacy in data fusion because homomorphic encryption is computationally expensive and may result in a higher error rate compared to differential privacy.

4.2 Applications of Data Fusion in Healthcare Settings

Combining data from various imaging modalities such as Magnetic Resonance Imaging (MRI), Computerized Tomography (CT), and Positron Emission Tomography (PET) improves diagnostic accuracy. Data fusion enables a more comprehensive understanding of a patient's condition. Integrating information from different EHRs creates a more complete patient profile, aiding in better diagnoses and treatment planning.

Challenges: Ensuring interoperability and standardization of data formats across different healthcare systems. Fusion of data from fitness trackers and smartwatches with other health information provides a holistic view of a patient's health, helping identify potential risks. Combining genetic data with other health information enhances the understanding of a patient's disease risk and supports personalized treatment plans.

4.3 Challenges Encountered in Healthcare Data Fusion

Diverse data sources may have varying levels of quality, formats, and standards, making it challenging to harmonize and standardize for meaningful fusion. Healthcare data is highly sensitive, and data fusion requires robust privacy measures to prevent unauthorized access. Compliance with regulations like Health Insurance Portability and Accountability Act (HIPAA) is crucial. Ensuring the seamless integration of data from different healthcare systems and devices requires addressing interoperability challenges, including data format and protocol differences. Data fusion techniques, especially those involving encryption and homomorphic operations, can be computationally intensive, affecting system performance and response times. Balancing the benefits of data fusion with ethical considerations, such as patient consent, transparency, and the responsible use of healthcare data is crucial. Incomplete data from certain sources or missing values can affect the accuracy of the fused data. Developing robust imputation methods is essential. Handling the exponential growth of healthcare data, especially in genomic and medical imaging applications, requires scalable data fusion methods and infrastructure. Adhering to regulatory frameworks and standards, such as GDPR in Europe or local healthcare regulations, is crucial to ensure legal compliance and patient trust.

4.4 Discussion:

Subcellular Location Dataset contains information about the subcellular localization of proteins. Subcellular localization refers to the specific compartment or organelle within a cell where a protein is found to be located. This dataset may include details such as the name of the protein, its corresponding subcellular location, and additional information such as experimental evidence supporting the localization. The RNA Cell Line Dataset contains RNA expression data across different cell lines. RNA expression data provides insights into the level of gene

expression in various cell types or conditions. This dataset includes information such as the gene name, cell line identifier, and the expression level of the corresponding gene in each cell line. Table 3 presents sample encryption parameters used in the Paillier cryptosystem, as well as the resulting fused ciphertexts obtained from secure data fusion. It includes the public and private keys, the original feature, the encrypted feature, the decrypted feature, and the ciphertext from the two data sources before fusion.

5. Conclusion

The proposed work demonstrates how to perform secure data fusion using homomorphic encryption with the Paillier cryptosystem. It generates a pair of public and private keys, encrypts data from two sources using the public key, performs homomorphic addition to fuse the encrypted data, and then decrypts the fused data using the private key. The resulting fused data is stored in the database for analysis and diagnosis. This helps in developing customized medical treatments. This technique can be useful in scenarios where multiple parties need to share data to perform a computation but are unable or unwilling to share the raw data due to privacy or ownership concerns. The use of homomorphic encryption ensures that the computations can be performed securely without any party seeing the other's data.

Using homomorphic encryption, specifically the Paillier cryptosystem, for secure data fusion can provide a way to combine data from multiple sources while ensuring privacy and security. However, the execution time and memory usage of the data-level Paillier cryptosystem approach may be higher than the feature-level fusion method due to the computationally intensive operations involved. Therefore, the choice of technique depends on the specific requirements of the use case, and a thorough analysis of the advantages and disadvantages of each technique should be conducted before choosing one over the other. Scalability can also be a challenge, but parallelizing encryption and decryption steps, optimizing the homomorphic addition algorithm or

Table 3 Encryption Parameters and Fused Data

Input	**Parameter**	**Value**
	Public Key (N, gen)	(11894275670047118937561259855360584038660241016795964172911746714981389696268, 45326375740269962684688010045818245521544802239683142713280284491108671555162)
	Private Key (λ, μ)	(39647585566823729791870866184535280128053721645055342084651478017924785129 16, 11174967258269649336888415408570734798498333202381563837473783849416660315032)
Output	**Feature**	**10**
	Encrypted Feature	53552272584278445582460112136313268001400298699832221840571788362856540596067507171268757600261786514087811860114898072913714196051932926030801421394784
	Encrypted Feature	53552272584278445582460112136313268001400298699832221840571788362856540596067507171268757600261786514087811860114898072913714196051932926030801421394784
	Decrypted Feature	8979697472691469644331087327684221955423406574983158684188650099758289812858
	Ciphertext 1	43723519548643567111546670557816785247776158678117489722511901027969866537454933730783317015456125520459151892426736937275293670216666435658872545305392
	Ciphertext 2	63464899467876628689732607347288501733742689873652280536125404402732109051181301528006935028439268162895452734809588352480999100557253891848777812340768
	Fused Ciphertext	10718841901652019580127927790510528698151884855176977025863730543070197558863623525879025204389539368335460462723632528975629277077392032750765035764616 0

using more efficient cryptographic primitives can improve it. Differential privacy is another technique used for preserving privacy in data analysis, offering a more lightweight approach than the Paillier cryptosystem, and its choice depends on the specific requirements of the use case. A secure data fusion block diagram aims to ensure that data from multiple sources is combined in a secure and reliable way, while also protecting the confidentiality and integrity of the data.

6. Acknowledgements

Shobha designed, analyzed, and interpreted the data regarding Gene clustering. Dr NN was a major contributor in writing the manuscript. All authors read and agreed on the final manuscript. We thankfully acknowledge Nitte Meenakshi Institute of Technology, Bengaluru, Affiliated to Visvesvaraya Technological University (VTU), Belagavi, Karnataka 590018 for constant support.

7. References

Albahri, A. S., Duhaim, A. M., Fadhel, M. A., Alnoor, A., Baqer, N. S., Alzubaidi, L., ... & Deveci, M. (2023). A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion. *Information Fusion*, *96*, 156-191. https://doi.org/10.1016/j.inffus.2023.03.008

Al-Hawawreh, M., & Hossain, M. S. (2023). A privacy-aware framework for detecting cyber attacks on internet of medical things systems using data fusion and quantum deep learning. *Information Fusion*, *99*, Article 101889. https://doi.org/10.1016/j.inffus.2023.101889

Alipour, M., La Puma, I., Picotte, J., Shamsaei, K., Rowell, E., Watts, A., ... & Taciroglu, E. (2023). A multimodal data fusion and deep learning framework for large-scale wildfire surface fuel mapping. *Fire*, *6*(2), Article 36. https://doi.org/10.3390/fire6020036

Anita, J. N., & Kumaran, S. (2023). A Systematic Review for Medical Data Fusion Over Wireless Multimedia Sensor Networks. *Artificial Intelligence for Sustainable Applications*, 117-126. https://doi.org/10.1002/9781394175253.ch7

Babu, B. S., & Narayana, M. V. (2023). Two stage multi-modal medical image fusion with marine predator algorithm-based cascaded optimal DTCWT and NSST with deep learning. *Biomedical Signal Processing and Control*, *85*, Article 104921, https://doi.org/10.1016/j.bspc.2023.104921

Dai, Y., Yan, Z., Cheng, J., Duan, X., & Wang, G. (2023). Analysis of multimodal data fusion from an information theory perspective. *Information Sciences*, *623*, 164-183. https://doi.org/10.1016/j.ins.2022.12.014

Ifzarne, S., Hafidi, I., & Idrissi, N. (2023). Compressive sensing and paillier cryptosystem based secure data collection in WSN. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), 6243-6250, https://doi.org/10.1007/s12652-021-03449-6

Jemili, F. (2023). Towards data fusion-based big data analytics for intrusion detection. *Journal of Information and Telecommunication*, *7*(4), 409-436. https://doi.org/10.1080/24751839.2023.2214976

Liang, Z., Zhang, Z., Chen, H., & Zhang, Z. (2022). Disease prediction based on multi-type data fusion from Chinese electronic health record. *Mathematical Biosciences and Engineering*, *19*(12), 13732-13746. https://doi.org/10.3934/mbe.2022640

Mergin, A., & Premi, M. G. (2023). Shearlet transform-based novel method for multimodality medical image fusion using deep learning. *International Journal of Computational Intelligence and Applications*, *22*(01), Article 2341006. https://doi.org/10.1142/S1469026823410067

Mohsen, F., Ali, H., El Hajj, N., & Shah, Z. (2022). Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Scientific Reports*, *12*(1), Article 17981, https://doi.org/10.1038/s41598-022-22514-4

Nguyen, C., Sagan, V., Bhadra, S., & Moose, S. (2023). UAV Multisensory Data Fusion and Multi-Task Deep Learning for High-Throughput Maize Phenotyping. *Sensors*, *23*(4), Article 1827, https://doi.org/10.3390/s23041827

Qi, L., Chi, X., Zhou, X., Liu, Q., Dai, F., Xu, X., & Zhang, X. (2022, August 22-25). *Privacy-aware data fusion and prediction for smart city services in edge computing environment* [Conference presentation]. 2022 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE

Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics). IEEE. Espoo, Finland.

Rainio, O., Han, C., Teuho, J., Nesterov, S. V., Oikonen, V., Piirola, S., ... & Klén, R. (2023). Carimas: an extensive medical imaging data processing tool for research. *Journal of Digital Imaging*, *36*(4), 1885-1893. https://doi.org/10.1007/s10278-023-00812-1

Raman, S. R., Qualls, L. G., Hammill, B. G., Nelson, A. J., Nilles, E. K., Marsolo, K., & O'Brien, E. C. (2023). Optimizing data integration in trials that use EHR data: lessons learned from a multi-center randomized clinical trial. *Trials*, *24*(1), Article 566. https://doi.org/10.1186/s13063-023-07563-y

Steyaert, S., Pizurica, M., Nagaraj, D., Khandelwal, P., Hernandez-Boussard, T., Gentles, A. J., & Gevaert, O. (2023). Multimodal data fusion for cancer biomarker discovery with deep learning. *Nature Machine Intelligence*, *5*(4), 351-362. https://doi.org/10.1038/s42256-023-00633-5

Singh, K. K., & Barde, S. (2024). A feasible adaptive fuzzy genetic technique for face, fingerprint, and palmprint based multimodal biometrics systems. *Journal of Current Science and Technology*, *14*(1), Article 1. https://doi.org/10.59796/jcst.V14N1.2024.1

Sunitha, T.O., Rajalakshmi, R. & Sujatha, S. S. (2022). Fuzzy based dynamic histogram equalization for enhancing quality of registered medical image. *Journal of Current Science and Technology, 12*(2), 243-264.

Tsai, C. S., Zhang, Y. S., & Weng, C. Y. (2022). Separable reversible data hiding in encrypted images based on paillier cryptosystem. *Multimedia Tools and Applications*, *81*(13), 18807-18827, https://doi.org/10.1007/s11042-022-12684-8

Zhang, J., Huang, Q., Huang, Y., Ding, Q., & Tsai, P. W. (2023a). DP-TrajGAN: A privacy-aware trajectory generation model with differential privacy. *Future Generation Computer Systems*, *142*, 25-40. https://doi.org/10.1016/j.future.2022.12.027

Zhang, P., Cheng, X., Su, S., & Wang, N. (2023b). Effective truth discovery under local differential privacy by leveraging noise-aware probabilistic estimation and fusion. *Knowledge-Based Systems*, *261*, Article 110213. https://doi.org/10.1016/j.knosys.2022.110213