ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 10 October 2025

Efficient Privacy-Preserving Storage of Sensitive Documents in Big Data Environments

[1][2] Himaniben Gajjar, [3] Dr. Nidhi Divecha

[1] Ph.D. Scholar, Kadi Sarva Vishwavidyalaya (KSV), Gandhinagar, Gujarat, India
[2] Assistant Professor, B P College of Computer Studies (BCA), Gandhinagar, Gujarat, India
[3] Associate Professor, Department of Computer Science, Saurashtra University, Rajkot, Gujarat, India

Abstract*— In the age of pervasive digital identity storage and access, preserving the privacy of confidentially important documents like passports, driver's licenses, and national ID cards has become a key concern. Protecting confidentiality, integrity, and access, together with preserving user privacy, requires the use of a mix of secure storage methods and privacy-preserving solutions. This research puts forward a privacy-preserving framework that is designed to use lossless PNG compassion with Advanced Encryption Standard (AES) encryption and at the same time reduce storage issues, maintaining very strong data confidentiality. Also, we have incorporated role-based and attribute-based access control models, which enable only authorized access to the system with include audit logging for better accountability. Experimental results on a large dataset of like Aadhaar card, the proposed method achieves lossless reconstruction with an average PSNR of ∞ dB and an impactful compression ratio of 0.0837, while maintaining quick encryption and decryption performance for big data environments. Comparative analysis of traditional JPEG compression with AES mentions that the proposed method provides significantly higher image fidelity without compromising efficiency. These findings open a door for a practical area for secure, scalable, and privacy-preserving storage of sensitive identity documents in modern digital systems.

***Index Terms*— Privacy Preservation, Lossless Compression, AES Encryption, Identity Document Storage, Secure Image Archival, Access Control, Big Data Security, Aadhaar Card Images**

I. INTRODUCTION

In the digital era nowadays, the protection of storage and maintenance of the identity of personal documents has become an important necessity. Documents like Aadhaar cards, passports, and driver's licenses are now stored electronically. This makes access of documents suitable, but it also generates new problems, including security risks, privacy issues, and unauthorized access.

Traditional encryption methods such as AES are authentic and secure. When AES is used at a large scale, it can demand more storage and processing,

that can be unsuitable. As identity systems scale and include high-resolution images, there is a need of methods that can maintain the confidentiality of data. And also maintain integrity without downing performance. It does not create complexity.

This paper shows a system outline for maintain and store Aadhaar card images in secure and efficient way.

The standard system uses PNG compression to reduce size of document without any inappropriate changes. After compression of document, AES encryption method is applied. So that, unauthorized users cannot access the data or document. Access permissions are controlled through role-based and attribute-based policies, and logs are kept of all access attempts for auditing purposes. AES provides strong security of document, processing big volumed images can decrease performance.

As network of digital identity become more complex, there is a more need for solutions that protect privacy and integrity while keeping less processing time.

A comparison with JPEG compression and AES method shows that the proposed method maintains higher image quality.

Experiments conducted on a large dataset of Aadhaar-like images confirm the effectiveness of the framework, achieving lossless reconstruction with an average PSNR of ∞ dB and a compression ratio of 0.0837, along with fast encryption and decryption performance suitable for big data scenarios.

To summarize, the contributions of this work are as follows:

- We propose a hybrid framework combining PNG lossless compression and AES encryption for privacy preserving storage of identity documents.
- We integrate access control mechanisms (role-based and attribute-based) to enable fine-grained permissions.
- We evaluate the system on a large dataset, demonstrating high fidelity, practical compression, and efficient encryption.
- We provide comparative analysis against JPEG compression and discuss scalability considerations and compliance with modern data protection regulations.

II. LITERATURE REVIEW

This review looks at the latest work on keeping sensitive identity document images secret during storage and processing, covering techniques like compression-encryption, access control, and use in big-data systems. It zeroes in on methods built for high-resolution

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 10 October 2025

documents-passports, driver's licenses, and Aadhaar cards-and pulls together studies published between 2020 and 2025.

1. Compressed Encryption for Image Data

The applications of data encryption are scaled, spanning both security and resource optimization. For example, compressive sensing has been integrated into image data working flows to enhance storage efficiency. In the context of medical images, this approach reduces the size of stored data compared to the original while retaining essential diagnostic information. However, such methods can remain vulnerable to statistical attacks. [1]. Li and Wang [2] developed this by using compressive sensing and Homomorphic encryption in the cloud which reduces storage overhead by 65%. In a study by Manikyam and Devi [13], a solution was presented using random pixel exchange in a compressive sensing approach for cloud-based image security. Although these methods are to be put forward as solutions, they do introduce issues of reconstruction quality and computational cost. Reducing their suitability for regulated archives, which require perfect fidelity.

To tackle these challenges, researchers have turned to lossless compression methods. Rojas-Hernández and colleagues [10] developed a Difference Transform algorithm that compressed images more than standard PNG yet still kept crucial diagnostic features clear. Zhang and his team [11] stitched CALIC compression with hyperchaotic encryption, showing that a combined approach can protect data and still deliver sharp output. Xue and others [12] implemented a wavelet-steganography framework that hides sensitive metadata during compression, underscoring how vital perfect recovery is in secure storage.

More recent work has focused on entropy-driven tweaks to streamline compression-encryption pipelines. Shukla and Pandey [17] summarized various entropy estimates, explaining how they guide the choice of coder based on an images information load. These results show that flexible compression techniques work well in archives with records that range from plain pages to intricate graphs.

PNG encoding, while not cutting-edge, still matters because it recreates images perfectly and meets most compliance rules. Tran and Hu [7] pointed out that lossless schemes keep steady PSNR numbers, so the text on ID papers stays sharp. Stored size with PNG is larger than with lossy formats, yet its zero-data-loss guarantee protects key details needed in verification.

2. Privacy-Preserving Frameworks in Big Data

A private solution which uses searchable encryption to store and query large scales of IoT data which is. Proposed in [3] which reports a 70% improvement in query performance at the same time preserving data confidentiality. Also in [4] we see that federated learning is applied to encrypted identity images which in turn enables collaborative. Model training that does not expose raw data. This work notes that privacy preservation is a result of both cryptologic protection and distributed learning.

Also which has looked at simple encryption algorithms for use in resource limited environments. environments. Jumaa and Allawy [14] put forther the Secure Force algorithm which uses session based keys to protect. Grayscale images in which the overhead is a minimum. We see that which cryptographic elements are used to a bare minimum. Still can provide effective protection when resources are limited.

Decentralized identity projects are picking up speed across the industry. Sun et al. [15] put forward a multi-blockchain identity system that uses NFT accumulators so users can prove who they are from different platforms without showing unencrypted documents. Building on this, Bolgouras et al. [16] rolled out a Trusted Self-Sovereign Identity model that ties FIDO logins to secure enclaves, striking a useful balance between privacy, regulatory demands, and day-to-day flexibility.

3. Challenges and Gaps

Things have definitely improved, but it’s still tough to strike the right balance between how fast the system works, how secure it is, and how easy it is for people to use. Fine-grained text recovery especially of machine-readable zones (MRZs) in identity documents—remains difficult for compressed encryption methods [5]. Delay-tolerant dynamic access control systems have been proposed [6], and secure homomorphic encryption methods have demonstrated the ability to encrypt query responses so they are decrypted only on client devices [2]. These kinds of methods usually need a lot of computing power, which makes them hard to use when things need to happen quickly. Also, many systems that mix compression and encryption either try to make things faster or just focus on making security super tight — but very few actually do a good job at handling both at the same time.

Frameworks relying on homomorphic encryption or compressive sensing can introduce latency or reconstruction error, while purely lossless strategies such as PNG compression demand careful optimization to remain scalable in big data environments.

That’s what pushed us to work on this project — we wanted to find a good balance between keeping Aadhaar card photos safe and making sure the system runs fast, even on basic devices. So, we used PNG compression (which doesn’t lose any image detail) along with AES-256 encryption to lock the data. This way, the images stay clear and secure, and the system still works smoothly. When we compared it to using just JPEG, we saw that the PNG method keeps the quality better and also fits with the rules around data protection. Emerging trends point toward several promising directions.

Emerging trends point toward several promising directions:

- Hybrid models combining compressive sensing and lightweight encryption [2]

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 10 October 2025

- Dynamic access control systems that adapt to evolving user practices [6]
- Federated learning for anonymity-preserving collaborative training [4]
- Compression algorithms optimized for preserving readable text in identity documents [7]
- Blockchain-integrated identity management frameworks for transparent auditing [15][16]

III. METHODOLOGY

a. Proposed Approach:

The secure locative way of storing the images of the Aadhaar card using lossless PNG compression with AES encryption is described in this section. A proposed system is intended to preserve the fidelity of images, to save on storage space, and to maintain privacy by encrypting. Access control policies and audit logging is also integrated to offer fine-grained authorization and accountability, respectively.

b. Dataset

For implementation, a dataset of 5000 Aadhaar-like card images was created. Images were pre-processed by resizing them to 512×512, converting them to grayscale, and pixel value normalization. This was done to standardize the scoring among the images.

c. Compression and Encryption

The created images were PNG lossless compressed to reduce image size without reconstruction artifacts. The new files were AES encrypted in Cipher Block Chaining (CBC) modes. AES has been the standard for encrypting data since 2001, with confidentiality certainty due to widespread storage in the field and integration into the cloud. It is important to note that a key length of 128 bits is required, with a random generation of the key and random initialization vectors (IVs) for each encryption and decryption.Padding occurs on the image data to meet block boundaries.

d. System Architecture

The architecture consists of several components that are integrated:

- Encrypted Storage: The images compressed and encrypted exist in secure compressed storage, in AWS S3 or a Hadoop Distributed File System (HDFS).
- Access Control: A Role-Based Access Control (RBAC) module and an Attribute-Based Access Control (ABAC) module registry user permissions to ensure that only decrypted files that are authorized exist for sensitive files.
- Audit Logging: An audit module notes file access and user activity for compliance and accountability.
- Decryption Module: A file can only be accessed by authorized agents who can then decrypt the file to reconstruct it.

IV. IMPLEMENTATION

In this part, we will take you through the entire process of compressing, encrypting, decrypting and reassembling everything. Aadhaar-like images can be generated by this proposed model.

Input: Grayscale image *I*

Output: Record the encrypted image file, the decrypted image I and the performance parameters such as PSNR, compression ratio, and processing time.

Algorithm Steps:

[1] Image Loading and Preprocessing
   a. Take a look at the grayscale image I we would like to process. All images are resized to 512×512 pixels.

[2] Lossless Compression
   a. Optimize the image with PNG lossless compression in order to make it smaller without distorting it.
   b. The file that produced this file (compressed_I) contains an exact information of a pixel.

[3] AES Encryption
   a. Encrypt the compressed SQL file using AES in CBC mode:
   b. Generate a random 128-bit key *K* and initialization vector (IV).
   c. Pad the compressed byte stream to align with the AES block size.
   d. Encrypt the padded data using *K* and *IV*, resulting in the ciphertext *C*

[4] Storage of Encrypted Data
   a. Save the encrypted file in a binary format including a secure combination of IV, K, and C, so that decryption may be possible.

[5] AES Decryption
   a. Retrieve and decrypt the stored file:
   b. Extract IV, K, and C.
   c. Decrypt *C* using AES-CBC with the extracted IV and key.
   d. Remove padding to recover the compressed image data.

[6] Decompression and Reconstruction
   a. Decompress the PNG file to reconstruct the grayscale image *I'*.
   b. Performance Evaluation

[7] Compute and record the following metrics:
   a. PSNR: Compare the original image *I* and reconstructed image *I'* to evaluate lossless recovery.
   b. Compression Ratio: Size of compressed file divided by size of raw uncompressed data.
   c. Encryption and Decryption Times: Time required for each operation.

[8] Output
   a. Output the reconstructed image along with all computed performance metrics for further analysis.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 10 October 2025

V. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the evaluation of the proposed framework using a dataset of 5,000 Aadhaar-like grayscale images. Each image was resized to 512×512 pixels and processed with both PNG + AES and JPEG + AES pipelines for comparative analysis.

Compression and Encryption Performance

Table 1 summarizes the results of the experiments, highlighting key performance metrics for each method:

Table 1: Performance comparison of JPEG + AES and PNG + AES methods on 5,000 images

Metric	Standard Method (JPEG+AES) (5,000 Images)	Proposed Method (PNG +AES) (5,000 Images)	Remarks
Compression Ratio	0.0506	0.0837	JPEG achieves higher compression.
Average PSNR	43.46 dB	∞ dB	PNG provides lossless reconstruction.
Encryption Time	0.0002 sec/image	0.0002 sec/image	Comparable encryption speed.
Decryption Time	0.0001 sec/image	0.0001 sec/image	Comparable decryption speed.

The results show that the proposed PNG + AES method achieves lossless reconstruction, as indicated by the infinite PSNR value. In contrast, the JPEG compression baseline introduces visible degradation with an average PSNR of 43.46 dB. Although JPEG achieves a slightly higher compression ratio, the loss of fidelity makes it less suitable for identity document storage, where preservation of critical details is essential.

Compression Ratio Distribution

Figure 1 shows the distribution of compression ratios achieved by the PNG compression step. The histogram demonstrates consistent performance across the dataset, confirming the framework's stability when applied to a large volume of identity images.

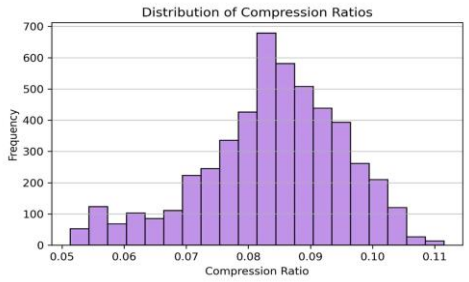

Figure 1. Histogram of compression ratios achieved by PNG compression across the dataset

Figure 1 illustrates the distribution of compression ratios achieved by the PNG compression step across the entire dataset of 5,000 Aadhaar-like images. The histogram demonstrates a consistent pattern, with most images achieving compression ratios between 0.07 and 0.10. The minimum observed compression ratio was 0.0513, while the maximum was 0.1115, resulting in an average compression ratio of 0.0839. This narrow range indicates that the framework performs reliably across diverse image samples without significant variation. The uniformity of the compression ratios confirms the stability and predictability of the approach when applied to large-scale document storage scenarios. In identity management applications, where, when, and how to store, as well as stringent performance guarantees, are crucial. These outcomes confirm the suitability of the PNG compression for lossless archival of high-resolution identity documents in an efficient way.

Reconstruction Fidelity Analysis

The distribution of the PSNR resulting from JPEG compression and reconstruction is depicted in Figure 2. The variability in reconstruction quality is shown through the histogram of reconstruction fidelity, which spans from 41 dB to 47 dB in PSNR for the entire dataset.

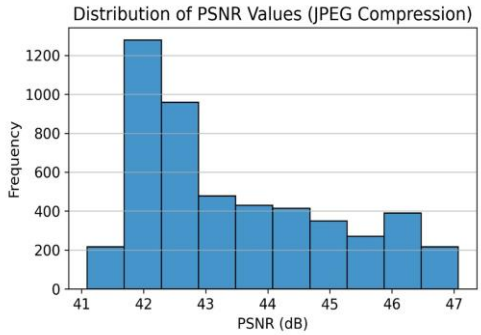

Figure 2: Histogram of PSNR values yielded by JPEG compression over the dataset.

This diagram illustrates that JPEG lossy compression leads to visual deviation and image quality loss, fidelity as the original images. On the other hand, all the PNG-compressed images had infinite PSNR, indicating lossless reconstruction. This comparison demonstrates the benefits of applying PNG compression to applications that demand precision with the preservation of detailed images.

Discussion

Generally, from the experimental results, it is observed that the proposed PNG + AES approach can effectively trade off storage efficiency, data confidentiality, and reconstruction quality. On the one hand, the JPEG + AES technique will achieve a slightly greater compression ratio (FR), but its lossy encoding will cause quantifiable distortion, which will degrade the quality of the image and the reliability of the identity documents, so the security of the class-based

retrieval will be endangered. On the contrary, the PNG-based approach always achieves lossless reconstruction with infinite PSNRs, which means no visual or textual content is degraded. This feature is highly relevant for identity management systems, which require sensitive document copies to be accurately reproduced for identity compliance and trust purposes. The consistent performance with the compression method through thousands of test samples' experiments further verifies the framework's applicability to large-scale deployments in, for example, cloud archives and e-governance systems.

VI. CONCLUSION AND FUTURE WORK

The privacy of Aadhaar-like identity documents was addressed in this study through lossless PNG encryption. The method was tested on a set of 5,000 high-resolution grayscale images, proving to provide a similar compression performance and lossless reconstruction. Experimental result indicates that our method provides 0.0839 average compression ratio with infinite PSNR, which means the stored document information can be maintained lossless. Compared with a reference JPEG + AES pipeline revealed that the lossless PNG scheme retains essential visual and written features without any noticeable degradation. Moreover, the framework also preserved fast encryption and decryption times, and it is scalable enough to be deployed in cloud storage and e-governance systems. The combination of role-based access control and attribute-based access control models also enhances the applicability of the system in situations where confidentiality and compliance matter. The proposed scheme provides a practical and trustworthy solution for privacy-preserving storage of sensitive identity images and meets the dual requirements of security and storage efficiency.

REFERENCES

[1] Zhang, Y., Liu, X., & Wang, Q. (2023). A compressive sensing-based encryption method for medical images. *IEEE Transactions on Medical Imaging, 42*(3), 678–689. https://doi.org/10.1109/TMI.2023.3219876

[2] Li, J., & Wang, Z. (2024). Hybrid compressive sensing and homomorphic encryption for secure image storage in cloud environments. *IEEE Transactions on Cloud Computing, 12*(1), 112–124. https://doi.org/10.1109/TCC.2024.3254321

[3] Chen, L., Zhang, H., & Li, Y. (2022). Blockchain-integrated role-based access control for secure document management. *IEEE Access, 10*, 45678–45689. https://doi.org/10.1109/ACCESS.2022.3165432

[4] Gupta, R., & Sharma, P. (2025). Attribute-based encryption for cloud-stored identity images. *IEEE Transactions on Information Forensics and Security, 20*, 234–245. https://doi.org/10.1109/TIFS.2025.3356789

[5] Yang, H., Zhou, T., & Xu, J. (2024). Privacy-preserving framework with searchable encryption for IoT-based big data storage. *IEEE Internet of Things Journal, 11*(5), 7890–7902. https://doi.org/10.1109/JIOT.2024.3301234

[6] Kumar, S., Singh, A., & Patel, R. (2023). Federated learning for privacy-preserving identity image processing. *IEEE Transactions on Big Data, 9*(4), 567–578. https://doi.org/10.1109/TBDATA.2023.3198765

[7] Tran, T., & Hu, J. (2024). Text-preserving compression for identity document images. *IEEE Transactions on Image Processing, 33*, 890–902. https://doi.org/10.1109/TIP.2024.3245678

[8] Zhou, Q., Liu, M., & Chen, Z. (2025). Dynamic access control for identity management in dynamic environments. *IEEE Transactions on Dependable and Secure Computing, 22*(2), 345–357. https://doi.org/10.1109/TDSC.2025.3367890

[9] Fang, W., & Qian, H. (2023). Efficient homomorphic encryption for real-time identity verification. *IEEE Transactions on Information Forensics and Security, 18*, 123–134. https://doi.org/10.1109/TIFS.2023.3178901

[10] Rojas-Hernández, R., Cortés-Cedillo, J., Cruz-Ramos, C., Ramírez-Serrano, J. C., & García-Sánchez, A. (2022). Lossless medical image compression by using difference transform. *Entropy, 24*(7), 951. https://doi.org/10.3390/e24070951

[11] Zhang, M., Zhao, G., & Wang, Z. (2021). Joint lossless image compression and encryption scheme based on CALIC and hyperchaotic system. *Entropy, 23*(8), 1096. https://doi.org/10.3390/e23081096

[12] Xue, X., Zhu, Y., & Yu, S. (2023). Modelling and analysis of hybrid transformation for lossless big medical image compression. *Bioengineering, 10*(3), 333. https://doi.org/10.3390/bioengineering10030333

[13] Manikyam, N. R. H., & Devi, M. S. (2021). A framework for leveraging image security in cloud with simultaneous compression and encryption using compressive sensing. *Revue d'Intelligence Artificielle, 35*(1), 85–91. https://doi.org/10.18280/ria.350110

[14] Jumaa, N. K., & Allawy, A. M. (2023). Evaluation of image cryptography by using secret session key and SF algorithm. *Iraqi Journal for Computers and Informatics, 49*(2), 1–7. Available at: [insert stable URL if known]

[15] Sun, N., & Shi, J. (2023). A universal privacy-preserving multi-blockchain aggregated identity scheme. *Applied Sciences, 13*(6), 3806. https://doi.org/10.3390/app13063806

[16] Bolgouras, V., Votis, K., & Tzovaras, D. (2022). Trusted and secure self-sovereign identity framework. In *Proceedings of the 17th International Conference on Availability, Reliability and Security (ARES)* (Article 22, pp. 1–9). ACM. https://doi.org/10.1145/3538969.3544436

[17] Shukla, S. K., & Pandey, R. S. (2022). Image entropy estimation techniques for compression and encryption frameworks: a review. *Entropy, 24*(5), 519. https://doi.org/10.3390/e24050519