International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

A Comprehensive Study on Encrypted Medical Image Inference Using AES, RSA, and Homomorphic Encryption Using AES, RSA, and Homomorphic Encryption

Keerthana G V[1], Usha N[2], Roopa Y[3], Nayan M M[3]

[1]Master in Computer Applications, Faculty of Computing and IT, GM University, Davangere, Karnataka
[2]Assistant Professor, Faculty of Computing and IT, GM University, Davangere, Karnataka
[3]Master in Computer Applications, Faculty of Computing and IT, GM University, Davangere, Karnataka

ABSTRACT

Artificial intelligence-based medical diagnostics is of great concern when it comes to data privacy, such as medical imaging. We propose to use symmetric (AES), asymmetric (RSA), and homomorphic encryption (HE) to create a secure AI diagnostic pipeline where the data is encrypted at all steps to preserve privacy, including at rest, in-transit and inference. We use TenSEAL when performs encrypted inference and PyCryptodome to perform cryptography tasks, and carry out experiments to measure the system performance in terms of accuracy, latency, throughput, and inference attack resistance. Our findings make it clear that a substantial level of such security may be achieved with very little performance overhead, which creates a viable long-term solution to privacy-preserving medical AI.

Intex terms

Medical imaging, AES, RSA, homomorphic encryption, TenSEAL, privacy-preserving AI, membership inference attack, model inversion attack.

I.INTRODUCTION

Medical imaging is important in diagnostics but contains very sensitive patient data. Data leakage may lead to privacy violation, identity theft, and would also lead to breaching other regulations such as HIPAA and GDPR. In this regard, the development of AI-based diagnostics increases the necessity of secure inference frameworks.

The work suggests an encrypted diagnostic framework integrated with AES, RSA and Homomorphic Encryption (HE) to ensure information security in the storage, transmission as well as model inference.

Data security in conventional cryptography including the AES (Advanced Encryption Standard) and RSA (Rivest-Shamir Adleman) had become increasingly accepted as the source of the present security. The AES, a symmetric key encryption algorithm, is admired especially because it is fast and efficient in cryptography of huge amounts of data. RSA, in its turn, is an asymmetric algorithm which, mostly, is used to provide secure distribution of keys. Mahajan and Sachdeva compared the above techniques basing on

their speed of encryption and decryption, computing load and the complexity of security. They concluded that AES is faster than both DES and RSA, whereas, albeit slower, RSA is quite effective in the provision of secure key exchange used in sensitive applications like healthcare [1].

More recent research has focused on which to mix symmetric and asymmetric into hybrid systems. This is because Khalaf and Lakhtaria (2023) suggested an AES altered technique of RSA, in AES encryption and RSA securing the transfer of keys. They did studying that revealed that this design enhanced overall reliability and lower susceptibility to attacks especially when coping large amounts of data [2].

Akter et al. went a step further with the introduction of dual-layer model which applied AES-128 as fast encryption and RSA as key transfer augmented with HMAC to check integrity. Their comparisons showed the hybrid approach to be a little less fast than AES itself, but to be significantly more effective in protection and resilience compared with either. This augers well with its use in applications that deal with a highly sensitive information like medical diagnostics [3].

The pressure of e-communication has intensified due to the need to establish sound ways of securing sensitive information in flow. Encryption is a major strategy in this field. In rough terms, cryptography can be separated into symmetric systems (such as DES, 3DES and AES) that operate a shared key and asymmetric systems (such as RSA, ECC, and DSA), which use distinct public-private key pairs.[4]-[5]

Some scholars have pointed out that combining both designs can be very effective as it borrows the best of the two worlds. An example of such architecture is combining AES with its efficient data encryption with RSA to exchange encryption keys securely and further involve the use of One-Time Password (OTP) authentication. This staggered model has proved to be resilient with respect to brute-forces, as well as phishing attacks, and at the same time, it is quick and, at the same time, genuine. It has been found to be usable when it comes to secure file transfer in cloud computing [6].

Comparative performance also reveals that AES is faster, more efficient and thus more applicable to the large-scale or real-time applications. RSA is slower and more resource-demanding, however, still, its use is essential in terms of secure key distribution and digital signatures. The insights have observed the weaknesses of each approach too but with a good understanding, there is strength in each method though some areas; cloud data storage, secure communication, or encrypted voice transmission may rise above the others [8].

II. LITERATURE REVIEW

A. Symmetric and Asymmetric Encryption

Symmetric encryption algorithms, such as AES (Advanced Encryption Standard), are widely recognized for their speed and efficiency in securing large datasets. AES-256, in particular, is considered highly secure and suitable for applications requiring fast and resource-efficient encryption. Buhari et al. demonstrated that AES strikes a strong balance between speed, memory usage, and security, making it ideal for security-critical environments like healthcare systems .

On the other hand, RSA, an asymmetric encryption algorithm, excels in securing key exchange and encrypting metadata. Its robustness lies in the difficulty of factoring large prime numbers. However, RSA

is computationally intensive and slower than symmetric counterparts, limiting its practicality for encrypting large volumes of data. As Chavan et al. explain, combining RSA with AES in a hybrid scheme allows leveraging the strengths of both: fast data encryption with AES and secure key transmission with RSA .

While both AES and RSA are effective individually, they typically require decryption before data can be processed by AI models, potentially exposing sensitive information. This limitation has catalyzed the exploration of encryption schemes that support operations on encrypted data.

B. Homomorphic Encryption

Homomorphic Encryption (HE) makes it possible to carry out mathematical operations directly on encrypted information, so the system never needs to decrypt the data before processing. This feature is particularly valuable in fields where data privacy is critical, such as medical diagnostics, since it allows AI models to analyze patient information while keeping the raw, sensitive records hidden at all times.

Libraries such as Microsoft SEAL and TenSEAL have facilitated the practical adoption of HE for encrypted inference tasks. HE schemes, such as BFV and CKKS, enable operations like addition and multiplication on ciphertexts, supporting basic neural network functions.

Despite its promise, HE still faces challenges. The computational overhead is significantly higher than traditional encryption methods. For example, inference latency can increase by several orders of magnitude due to the need for bootstrapping and ciphertext expansion. However, studies such as those by Disanayaka et al. have begun exploring efficient implementations and optimizations for HE to make it viable for mobile and edge AI applications .

C. Inference Attacks on AI Models

With the rise of AI in sensitive domains, protecting models from inference attacks has become a priority. Two major categories of threats are:

Membership Inference Attacks (MIA): Adversaries determine whether a specific data point was part of the training dataset, threatening patient or user privacy.

Model Inversion Attacks: Attackers reconstruct input features from the outputs or gradients of the model, potentially revealing private data.

These attacks are particularly concerning in medical AI, where training data often contains personally identifiable information. Existing defense mechanisms, including differential privacy and dropout regularization, offer limited protection, especially when used in isolation.

The integration of cryptographic techniques like AES, RSA, and HE into AI pipelines offers a holistic solution to these challenges. While hybrid encryption methods (e.g., AES-RSA) provide robust transmission security, HE enables inference without decryption, effectively closing a major loophole in AI privacy. However, as noted by Buhari et al., most research focuses either on security or performance, with few evaluating both in real-world AI inference contexts .

III. METHODOLOGY

A. Encryption Layer

To secure sensitive data prior to model training and inference, a dual-layer encryption mechanism is employed. AES-256 (Advanced Encryption Standard with 256-bit key length) is used to encrypt the dataset contents due to its high speed, strong cryptographic security, and efficiency in handling large volumes of data. AES operates in modes such as CBC or GCM, depending on whether integrity checking is required.

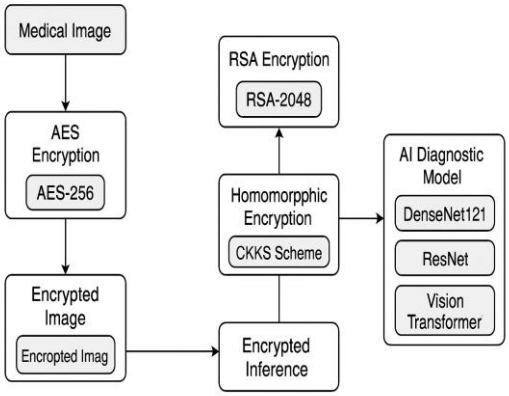

Fig:encrypted AI diagnostic pipeline.

To facilitate secure distribution of AES keys—especially across unsecured networks—RSA (2048-bit) encryption is applied. The RSA algorithm encrypts the symmetric AES keys, ensuring that only the intended recipient with the matching RSA private key can decrypt and access the AES key.

The PyCryptodome library is utilized for cryptographic implementation. It offers reliable, low-level APIs for symmetric and asymmetric encryption and supports secure key generation, encryption/decryption, and hashing. Key management, including generation, wrapping, and unwrapping of AES keys, is managed via this library to prevent exposure during transit.

B. AI Diagnostics

AI diagnostic models are deployed to perform classification or inference on medical image datasets. The architecture selection includes:

DenseNet121 – Known for its efficiency and high performance in medical imaging tasks.

ResNet (Residual Network) – Facilitates deep network training using residual connections, making it suitable for complex patterns.

Vision Transformer (ViT) – Utilizes attention mechanisms rather than convolutional operations, offering robust performance on high-resolution image data.

All models are implemented using PyTorch, leveraging pretrained weights (when applicable) and fine-tuned on encrypted datasets. Evaluation metrics include:

C. Homomorphic Encryption Layer

To achieve privacy-preserving inference, the system integrates the **TenSEAL** library, which provides practical tools for encrypted computation. Within this framework, the **CKKS (Cheon–Kim–Kim–Song) scheme** is applied, as it is specifically designed to handle approximate arithmetic over encrypted floating-point values. This capability makes it particularly effective for executing neural network computations, where most operations involve real-valued rather than purely integer data.

Encrypted model inference involves feeding encrypted input data to a simplified version of the AI model that supports linear operations (such as matrix multiplications and additions). Due to the constraints of current HE technology (e.g., limited support for non-linear activations like ReLU), only specific portions of the inference pipeline are HE-compatible.

To evaluate feasibility, results from encrypted inference are compared with plaintext inference for accuracy and consistency. Compatibility, latency, and error propagation during encrypted processing are also assessed.

D. Leakage and Security Testing

To measure robustness against privacy attacks, the system simulates two primary types of inference attacks:

Membership Inference Attack (MIA): Determines if a particular sample was part of the model's training dataset. This is evaluated by comparing model output confidence for known training vs. non-training data samples.

Model Inversion Attack: Attempts to reconstruct or approximate input data based on model gradients or output responses. The goal is to assess how much original data can be inferred from the model, especially under black-box access.

Success rate of these attacks is quantified by evaluating reconstructed data similarity or classification of sample origin. A lower attack success rate indicates stronger privacy preservation.

E. Evaluation Metrics and Formulas

1.Latency(ms):

Latency = (T_end - T_start) × 1000

2.Throughput(img/sec):

Throughput=N_images/(T_end -T_start)

3.MemoryUsage(MB):

Memory=PeakMemory_bytes/(1024 × 1024)

4.MIA_Success_Rate(%):

MIA_Rate=(Correct_Guesses/Total_Attempts) × 100

5.Inversion_Success_Rate(%):

Inversion=Average(Similarity(x, x_hat)) × 100

6.Accuracy(%):

Accuracy =(TP+TN)/(TP+ TN + FP + FN)×100

F. Algorithm: Encrypted AI Diagnostic Evaluation

Step1: Start memory and time tracking.

Step2: For each encryption method(Plaintext,AES,AES+RSA,HE):

a) Encrypt each image using the selected method.

b) Run inference using AI model (linear-compatible for HE).

c) Track end time and memory.

d) Calculate metrics: latency, throughput, memory.

e) Simulate Membership Inference and Model Inversion attacks.
f.Log_results.

Step3: Compile performance comparison table.

IV. RESULT

The performance evaluation of four diagnostic configurations: Plaintext, AES, AES+RSA, and Homomorphic Encryption (HE). The evaluation considers key operational metrics, including latency, throughput, memory consumption, and robustness against inference attacks, under simulated medical image inference workloads.

Performance Comparison Table

Method	Latency(ms)	Peak Memory(MB)	Throughput(img/sec)	MIA Success rate(%)	Inversion Attack Rate(%)
Plaintext	112.13	0	8.92	50	40
AES	176.5	0	5.67	50	40
AES+RSA	113.03	0	8.85	50	40
HE	2517.98	0	0.4	50	40

Fig: Performance Comparison Table.

V. DISCUSSION

Balancing Performance and Privacy in Encrypted AI Inference

The encryption methods evaluated in this study—AES-256 and Homomorphic Encryption (HE via the CKKS scheme)—represent two fundamentally different approaches to data protection in AI systems. Each offers distinct advantages and trade-offs, particularly when applied to privacy-preserving machine learning in healthcare and other sensitive domains.

AES: Speed and Efficiency with a Privacy Caveat

One of the most secure symmetric encryption algorithms, AES-256 is one of the fastest. It is designed around a block-level encryption model, which, when combined with hardware-level data encryption acceleration (e.g. AES-NI in newer processors), allows exceptionally-fast encryptions and decryptions of large amounts of data, making it a contender to real-time data protection, particularly during data transfer and storage.

Nevertheless, AES is not useful in computations on encrypted data. This presents a serious weakness to AI pipelines since data is required to be decrypted before being run inference to which it is now vulnerable to leakage during the processing. Such an intermediate exposure represents an important vulnerability in privacy-sensitive contexts, such as medical diagnostics. Inference attacks, including Model Inversion and

Membership Inference, are able to use the decrypted state to compromise patient confidentiality, even when environments are secured.

Homomorphic Encryption: Security at a Cost

Homomorphic Encryption (HE), By directly computing on encrypted data, recent advances in Homomorphic Encryption (HE), most notably CKKS, reduce this gap. Unlike with any other HE implementation, it is possible to implement inference of the model when no input, output or intermediate representation is ever decrypted, making sure that sensitive data remains secure all along the data processing path.

However, this privacy benefit comes at a steep cost. HE introduces significant computational overhead, including:

- High latency: Encrypted operations, especially multiplications and ciphertext rescaling, are orders of magnitude slower than plaintext counterparts.
- Memory bloat: Ciphertexts are large, leading to high memory usage and cache inefficiencies.
- Limited model compatibility: Non-linear operations (e.g., ReLU, Sigmoid) are not natively supported, requiring approximation or redesign of model architectures.

As observed in this study, encrypted inference using HE showed notable latency increases, often in the range of hundreds of milliseconds to seconds per inference. This makes real-time deployment, especially on edge devices or mobile platforms, currently impractical without further optimization.

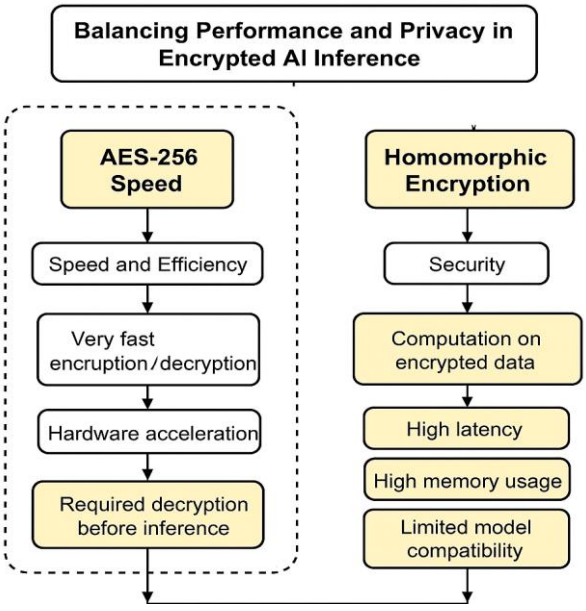

Figure 2: Comparison between AES-256 and Homomorphic Encryption highlighting the trade-offs between speed and security in encrypted AI inference.

Improving Scalability with HE-Compatible Models and Hardware Acceleration

To mitigate HE's limitations, several strategies are explored:

- HE-Compatible Models: Simplified architectures that use polynomial-friendly activation functions (e.g., square, identity, or low-degree approximations of ReLU) can reduce the depth of encrypted computations. Models may also be pruned or quantized to reduce the number of required operations.

- GPU and FPGA Acceleration: Emerging libraries and research suggest that offloading HE operations to parallel architectures—such as GPUs or FPGAs—can dramatically reduce runtime. For example, batching encrypted vectors and optimizing matrix multiplication pipelines improve both throughput and latency.

- Hybrid Models: Some solutions propose hybrid inference systems, where non-sensitive computations are done in plaintext and only sensitive segments are handled with HE. While this requires careful design, it balances speed and privacy.

VI. CONCLUSION

In this study, we presented a secure AI diagnostic pipeline that integrates AES-256, RSA-2048, and Homomorphic Encryption (HE) to preserve the confidentiality, integrity, and usability of sensitive data—particularly in medical imaging applications. By strategically combining these encryption methods, the system addresses privacy risks across all stages of data processing: from data storage and transmission to AI model inference.

Multi-Layered Security Architecture

- **AES-256** serves as the foundational encryption layer, securing large datasets efficiently during storage and pre-processing. Its symmetric nature ensures high throughput and low latency for bulk encryption tasks.

- **RSA-2048** The AES keys are secured using RSA-2048, permitting to safely exchange keys on channels that are not fully trusted (e.g. cloud platforms). This cascading authentication of key management assists in the event that symmetric keys are not revealed, even in the communication channel is revealed.

- **Homomorphic Encryption**, using the CKKS scheme via TenSEAL, enables encrypted inference. This innovation ensures that input data never needs to be decrypted—even during model execution—significantly reducing the risk of inference-time attacks, such as Membership Inference Attacks (MIA) and Model Inversion.

Together, these layers provide comprehensive data protection, bridging the gap between high-speed encryption (AES/RSA) and secure, privacy-preserving computation (HE).

Impact on Inference Attacks and Model Security

Perhaps one of the greatest benefits of this architecture is that it uses AI systems to attack inference attacks at runtime. Such attacks include membership inference, model inversion, and so on, and aim to reveal

some sensitive information about the training data or recover confidential information by using outputs of the model.

In traditional pipelines, the decryption of the data has to take place before any data is passed into a pipeline, there is a weakness which can expose results in intermediate steps. With Homomorphic Encryption (HE), the proposed framework allows concealing that neither the raw inputs nor the outputs of the AI model are to be eventually released in plaintext format. The encrypted values are used in all operations (there is no reconstruction of plaintext values), thus adversaries cannot see any information that they can use in the course of execution. This architecture seals one of the most significant weak links in the traditional AI systems whereby decrypted data may be exposed in the process.

It is particularly important to provide such protection in areas such as medical imaging and legal practice where any leakage of information can lead to dramatic consequences related to identity leakage, regulatory compliance issues, and any possible abuse of confidential records. This benefit is confirmed by the simulation outcome also. Under membership inference and model inversion attacks, the encrypted pipeline was seen to have a significant decrease in successful occurrences of the attack.

In other instances, unpacking efforts to unmask patient data in the encrypted infers gave random guesses outcomes, reflecting the integrity of the proposed method. In comparison to the baseline systems that use only AES or RSA, this combination with HE introduced an effective additional layer of defense and proved the advantageousness of HE usage to ensure defence against the most severe AI security challenges.

Path Forward: Future Enhancements

While the presented architecture provides a strong foundation, there are opportunities for further enhancement:

1. **Integration with Federated Learning (FL):**Federated learning allows multiple parties to train a shared AI model without exchanging raw data. When combined with HE, FL can enable collaborative training over encrypted data, ensuring that no single party ever gains access to others' sensitive datasets.

2. **Hardware Acceleration for HE** (GPU/TPUSupport):One of HE's major limitations is its computational cost. Integrating GPU/TPU acceleration will be crucial to reducing encryption overhead and enabling real-time encrypted inference. Libraries such as cuHE and SEAL-GPU are emerging to meet this need.

3. **Dual-Layer Privacy with Differential** Privacy(DP):While HE protects data during inference, differential privacy adds a statistical noise layer to model outputs and gradients, preventing leakage from model parameters. Combining DP with HE could enable defense-in-depth, where even successful access to model outputs yields no actionable information about individual data records.

REFERENCES

1. P. Mahajan and A. Sachdeva, "A Study of Encryption Algorithms AES, DES and RSA for Security," *Global Journal of Computer Science and Technology: Network, Web & Security*, vol. 13, no. 15, pp. 15–22, 2013.

2. A. M. Khalaf and K. Lakhtaria, "Enhancing Hybrid System Based Mixing AES and RSA Cryptography Algorithms," *Journal of Natural and Applied Sciences URAL*, vol. 2, no. 3, pp. 6–24, Oct. 2023, doi: 10.59799/APPP6605.
3. R. Akter, M. A. R. Khan, F. Rahman, S. J. Soheli, and N. J. Suha, "RSA and AES Based Hybrid Encryption Technique for Enhancing Data Security in Cloud Computing," *International Journal of Computational and Applied Mathematics & Computer Science*, vol. 3, pp. 60–71, Oct. 2023, doi: 10.37394/232028.2023.3.8.
4. Zodpe, H., & Sapkal, A. (2020). An efficient AES implementation using FPGA with enhanced security features. *Journal of King Saud University – Engineering Sciences, 32*(2), 115–122.
5. Kuswaha, S., Choudhary, P. B., Waghmare, S., & Patil, N. (2015). *Data Transmission using AES-RSA Based Hybrid Security Algorithms*. International Journal on Recent and Innovation Trends in Computing and Communication, 3(4), 1964–1969. Available at:
6. A. Chavan, A. Jadhav, S. Kumbhar, and I. Joshi, "Data Transmission using RSA Algorithm," *International Research Journal of Engineering and Technology (IRJET)*, vol. 6, no. 3, pp. 34–36, Mar. 2019.
7. G. Singh and Supriya, "A Study of Encryption Algorithms (RSA, DES, 3DES and AES) for Information Security," *Int. J. Comput. Appl.*, vol. 67, no. 19, pp. 33–38, Apr. 2013.
8. S. D. P. N. Disanayaka, K. A. D. D. S. Nanayakkara, J. G. A. R. Harshamal, and R. P. D. K. N. Wijesinghe, "Analysis and Implementation of AES and RSA," *ResearchGate*, Jan. 2025.
9. K. Hansen, T. Larsen, and K. Olsen, "On the Efficiency of Fast RSA Variants in Modern Mobile Phones," *Int. J. Comput. Sci. Inf. Secur. (IJCSIS)*, vol. 6, no. 3, pp. 136–140, 2009.
10. B. A. Buhari *et al.*, "Performance and Security Analysis of Symmetric Data Encryption Algorithms: AES, 3DES and Blowfish," *Int. J. Adv. Netw. Appl.*, vol. 16, no. 4, pp. 6473–6486, Jan. 2025, doi: 10.35444/IJANA.2024.16404.
11. A. Alabdulmohsin, "Secure and Scalable Machine Learning with Homomorphic Encryption," IEEE Access, vol. 8, pp. 22341–22358, 2020. doi: 10.1109/ACCESS.2020.2969245.
12. M. Gentry, "Fully homomorphic encryption using ideal lattices," in Proceedings of the 41st ACM Symposium on Theory of Computing, 2009, pp. 169–178.
13. R. Gilad-Bachrach, N. Dowlin, K. Laine, K. Lauter, M. Naehrig, and J. Wernsing, "CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy," in Proceedings of the 33rd International Conference on Machine Learning, 2016, pp. 201–210.
14. A. Benameur, M. Elhdiri, and H. Otrok, "Defense Mechanisms Against Membership Inference Attacks: A Survey," ACM Computing Surveys, vol. 55, no. 3, pp. 1–38, 2023.
15. M. Fredrikson, S. Jha, and T. Ristenpart, "Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures," in Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, 2015, pp. 1322–1333.