ViT Enhanced Privacy-Preserving Secure Medical Data Sharing and Classification

Al Amin*, Kamrul Hasan*, Sharif Ullah†, M. Shamim Hossain‡
*Department of Electrical & Computer Engineering, College of Engineering, Tennessee State University, TN, USA †Department of Computer Science & Engineering, College of Science and Engineering, University of Central Arkansas, Conway, AR, USA ‡Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
Email: {*aamin2, mhasan1*}@tnstate.edu, mullah@uca.edu, mshossain@ksu.edu.sa

2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC) | 979-8-3315-0805-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/CCNC54725.2025.10976131

Abstract—**Privacy-preserving and secure data sharing are critical for medical image analysis while maintaining accuracy and minimizing computational overhead are also crucial. Applying existing deep neural networks (DNNs) to encrypted medical data is not always easy and often compromises performance and security. To address these limitations, this research introduces a secure framework consisting of a *learnable encryption method based on block-pixel operation* to encrypt the data and subsequently integrate it with the *Vision Transformer (ViT)*. The proposed framework ensures data privacy and security by creating unique scrambling patterns per key, providing robust performance against leading bit attacks and minimum difference attacks.**

Index Terms—**Privacy-Preserving Data Sharing, Vision Transformer (ViT), Adversarial Robustness, Medical Image sharing and Classification.**

I. Introduction

Advancements in ML and DL models, supported by cloud platforms like Google Cloud and Microsoft Azure, offer high computational power but also introduce privacy risks in shared environments [1]–[3]. In response, the research proposed ViT enhances the learnable encryption method and obfuscates image details while preserving classification-relevant features, ensuring HIPAA compliance. When tested on MRI and histopathological datasets, the method showed strong defense against attacks such as leading bit-attack, and minimum difference attacks, keeping medical data safe in the cloud.

II. Proposed Approach

This research proposes ViT enhance privacy-preserving multi-layer learning approach for secure and robust medical image classification. Figure1 shows the overview of the approach. Multiple clients (C-1, C-2, ..., C-N) encrypt their local images using distinct keys (K-1, K-2, ..., K-N), applying a learnable encryption technique based on Block-Pixel Operation, including negative-positive transformation and color channel shuffling, to obfuscate image details while preserving essential features. These encrypted datasets are transmitted to a central server and processed through an Embedding Layer that converts the scrambled patches into a high-dimensional space for the Transformer. The Transformer Encoder then captures complex dependencies across the encrypted patches, enabling robust feature extraction, and a classifier produces accurate medical image classifications. The proposed framework bridges various data owners' diverse feature spaces and builds a more efficient global model. It also considers potential adversarial attempts to reconstruct the encrypted data to the original data, ensuring the security of sensitive medical data in cloud-based environments.

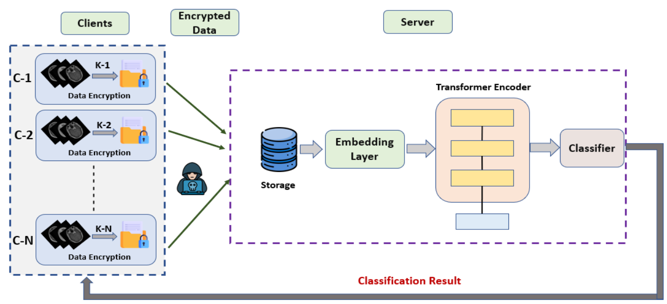

Fig. 1: Overview of ViT integrated privacy-preserving secure medical data sharing and classification framework.

A. Learnable Encryption based on Block-Pixel Operation

The Learnable Encryption technique secures medical images using block-wise scrambling, pixel and channel shuffling, and negative-positive transformations to obfuscate visual details effectively. The following section provides a mathematical representation of the components and operations of the proposed encryption system.

- **Block-Wise Scrambling:** Each patch P_i is subjected to a scrambling operation, where the pixels within the patch are reordered based on a key K. The scrambling operation is defined as:

$$S(P_i, K) = \text{Permute}(P_i, K) \tag{1}$$

  where $S(P_i, K)$ is the scrambled patch, and $\text{Permute}(P_i, K)$ is a function that reorders the pixels according to the key K.
- **Shuffling Block Positions:** After scrambling, the positions of the patches are shuffled based on the key K. Let the original patch positions be indexed as $\{1, 2, \ldots, 64\}$. The shuffling process is expressed as:

$$\text{Shuffle}(P_i, K) \rightarrow P_{\sigma(i)} \tag{2}$$

TABLE I: Performance Comparison of DNN Models and the Proposed Framework on Plain and Encrypted dataset

Plain Image	Encryption Method	Model	Dataset		Accuracy		Computational Time	
			Name	Type	training accuracy %	validation accuracy %	training time (m)	validation time (m)
	N/A	ResNet50	Brain Tumor	MRI	94	93	17.28	4.32
	N/A	MobileNet V2	Brain Tumor	MRI	98	96	31.62	7.9
Encrypted Image	pixel shuffling	ResNet50	Brain Tumor	MRI	46	38	541.6	108.32
		MobileNet V2	Brain Tumor	MRI	51	51	184.92	46.23
		Inception V3	Brain Tumor	MRI	48	46	432.36	86.46
	Proposed Encryption	**ViT (Proposed Work)**	**Brain Tumor**	**MRI**	**95**	**94**	**82.25**	**19.78**
			Lung and clone	Histopathological	95	94	98.11	24.53

where σ is a permutation of the patch indices determined by the key K.

- **Negative-Positive Transformation:** Each pixel value in the scrambled patches undergoes a conditional inversion, where the pixel value x is transformed as follows:

$$x' = \begin{cases} x & \text{if } r = 0 \\ 255 - x & \text{if } r = 1 \end{cases} \tag{3}$$

where r is a randomly generated binary integer (0 or 1) for each pixel, controlled by the key K.

- **Color Channel Shuffling:** The color channels of each pixel (R, G, B) are shuffled based on the key K. For a pixel with channels $C = [R, G, B]$, the shuffled channels C' are given by:

$$C' = [C_{\sigma(1)}, C_{\sigma(2)}, C_{\sigma(3)}] \tag{4}$$

where σ is a permutation of the channel indices $\{1, 2, 3\}$ determined by the key K.

- **Reconstruction of the Encrypted Image:** The final encrypted image $I_{\text{encrypted}}$ is reconstructed by assembling the scrambled and transformed patches:

$$I_{\text{encrypted}} = \text{Reassemble}(\{S(P_1, K), S(P_2, K), \ldots, S(P_{64}, K)\}) \tag{5}$$

where Reassemble refers to reassembling the full image from its encrypted patches.

B. Dataset Description and Model Performance

Two medical image datasets were used [4], 5,712 MRI images for brain tumor classification (with 1,311 images for testing) and 2,980 histopathological images for lung and colon cancer classification (with 311 for testing). As shown in Table I, the ViT model achieves higher training and validation accuracies of 95% and 94% on encrypted MRI brain tumor datasets, compared to the DNN models, which exhibit much lower accuracies ranging from 38% to 51%.

C. Security Evaluation

The proposed encryption method was rigorously evaluated against adversarial attacks, including leading bit attacks and minimum difference attacks. As illustrated in Figure 2, these attacks were unable to reconstruct the encrypted images, demonstrating the robustness of the model while maintaining high classification accuracy.

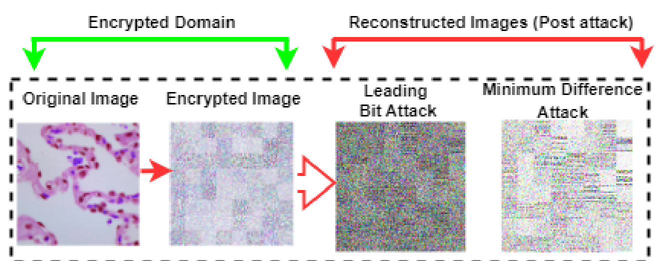

Fig. 2: Visual Representation of Transformation Pipeline: Original, Encrypted, and Post-Attack Images.

III. Conclusion and Future Work

This research introduces a secure Vision Transformer (ViT) model with learnable encryption for efficient medical image classification, outperforming traditional DNNs while protecting sensitive data. Future work will aim to scale the model, improve efficiency, and integrate federated learning for enhanced data privacy.

References

[1] K. Madono, M. Tanaka, M. Onishi, and T. Ogawa, "Block-wise scrambled image recognition using adaptation network," *CoRR*, vol. abs/2001.07761, 2020.

[2] X. Zhao, Z. Qi, S. Wang, Q. Wang, X. Wu, Y. Mao, and L. Zhang, "Rcps: Rectified contrastive pseudo supervision for semi-supervised medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 1, pp. 251–261, 2024.

[3] S. M. Nour, "Artificial intelligence (ai) for improving performance at the cutting edge of medical imaging," in *2023 5th Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, pp. 426–429, 2023.

[4] M. Nickparvar, "Brain tumor mri dataset," 2021.