scientific reports

Check for updates

OPEN

Deep learning steganography for big data security using squeeze and excitation with inception architectures

Bini M. Issac[1], S. N. Kumar[2]✉, Sherin Zafar[3]✉, Kashish Ara Shakil[4]✉ & Mudasir Ahmad Wani[5]

With the exponential growth of big data in domains such as telemedicine and digital forensics, the secure transmission of sensitive medical information has become a critical concern. Conventional steganographic methods often fail to maintain diagnostic integrity or exhibit robustness against noise and transformations. In this study, we propose a novel deep learning-based steganographic framework that combines Squeeze-and-Excitation (SE) blocks, Inception modules, and residual connections to address these challenges. The encoder integrates dilated convolutions and SE attention to embed secret medical images within natural cover images, while the decoder employs residual and multi-scale Inception-based feature extraction for accurate reconstruction. Designed for deployment on NVIDIA Jetson TX2, the model ensures real-time, low-power operation suitable for edge healthcare applications. Experimental evaluation on MRI and OCT datasets demonstrates the model's efficacy, achieving Peak Signal-to-Noise Ratio (PSNR) values of 39.02 and 38.75, and Structural Similarity Index (SSIM) values of 0.9757, confirming minimal visual distortion. This research contributes to advancing secure, high-capacity steganographic systems for practical use in privacy-sensitive environments.

Keywords Medical image security, Steganography, Squeeze-and-Excitation, Inception module, Deep learning, Jetson TX2, Telemedicine

In modern medical environments, the rapid growth of telemedicine and big data technologies has significantly transformed diagnostic workflows. For each patient, multiple scans across various imaging modalities and time points such as MRI, CT, and OCT are commonly acquired, resulting in high-resolution, large-scale datasets that contain highly sensitive medical information [1]. Following the COVID-19 pandemic, there has been an accelerated shift toward the use of cloud-based platforms for transmitting these medical images, enabling remote diagnosis and consultation. Telemedicine has proven instrumental in minimizing physical contact and maintaining continuity of care during times of healthcare system strain. However, the adoption of cloud-based remote diagnostics within distributed healthcare infrastructures also introduces critical challenges related to data confidentiality, integrity, and computational efficiency [2,3]. Consequently, ensuring secure and efficient transmission of medical images remains a paramount concern in contemporary healthcare systems [4].

Traditional steganographic techniques, such as Least Significant Bit (LSB) and Pixel Value Differencing (PVD), have served as baseline methods for embedding information within medical images. However, their limited resilience to noise, compression, and transformation renders them unsuitable for real-world applications in telehealth, where images are routinely subjected to varied transmission conditions. Moreover, these approaches often compromise the balance between embedding capacity and image quality—an unacceptable trade-off in diagnostic imaging contexts where visual fidelity is non-negotiable.

Recent advancements in Deep Learning (DL) offer promising avenues for adaptive and robust steganographic systems. Deep neural networks, particularly convolutional encoder-decoder architectures, have demonstrated

[1]Dept. of Computer Science & Engineering, Amal Jyothi College of Engineering, APJ Abdul Kalam Technological University, Thiruvananthapuram, Kerala 695 016, India. [2]Dept. of Electrical and Electronics Engineering, Amal Jyothi College of Engineering, Kanjirappally 686518, Kerala, India. [3]Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi, India. [4]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint AbdulRahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. [5]EIAS Data Science and Blockchain Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia. ✉email: appu123kumar@gmail.com; sherin.zafar@jamiahamdard.ac.in; kashakil@pnu.edu.sa

the ability to embed and reconstruct hidden data with high imperceptibility and robustness. Despite these capabilities, most current DL-based steganographic methods are computationally intensive and poorly suited for edge deployment, limiting their practical application in low-power medical environments.

To overcome these limitations, we propose a hybrid steganographic framework that integrates residual connections, Squeeze-and-Excitation (SE) attention mechanisms, and Inception modules to optimize both encoding and decoding performance. The proposed model is designed for real-time deployment on resource-constrained device NVIDIA Jetson TX2 platform, ensuring feasibility in edge computing environments. By using multi-scale feature extraction and channel-wise recalibration, our method enhances both the imperceptibility and robustness of medical image embedding.

The overall architecture of the proposed steganography framework is illustrated in Fig. 1. The system takes two input repositories: cover images dataset and secret medical images dataset. These datasets are divided into training and testing subsets to enable robust model development and evaluation. The encoder module comprises multiple 2D convolutional layers and Squeeze-and-Excitation (SE) blocks, which collaboratively extract and recalibrate essential features. These components ensure high-fidelity embedding while preserving the visual quality of the cover images. The output of the encoder is a set of stego images, which visually resemble the original cover images but contain the embedded secret information. The decoder receives these stego images and reconstructs the hidden medical data using a deep convolutional structure enhanced with residual connections and inception blocks. These layers help preserve spatial integrity and ensure accurate recovery of the embedded content. The trained model is deployed on the NVIDIA Jetson TX2 platform, a power-efficient edge computing device equipped with a CUDA-enabled GPU, which facilitates real-time inference and secure medical data transmission in low-resource settings. Also, performance evaluation is conducted using standard steganography metrics such as Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), Normalized Cross-Correlation (NCC), Average Difference (AD), Structural Similarity Index (SSIM) and Laplacian Mean Squared Error (LMSE) to assess image quality, embedding success, and reconstruction accuracy. The key contributions of our work are:

- A hybrid encoder-decoder architecture is implemented, integrating Squeeze-and-Excitation (SE) blocks, dilated convolutions, and Inception modules to enhance embedding robustness and reconstruction fidelity in medical image steganography.
- The decoder network employs residual connections and multi-scale feature extraction to recover secret images with minimal distortion under challenging conditions (e.g., noise, compression).
- Our model is optimized for edge deployment on the NVIDIA Jetson TX2 platform, enabling real-time, low-power operation for telemedicine and teleradiology.
- Extensive experiments conducted on Brain MRI and OCT glaucoma datasets demonstrate superior performance in terms of PSNR, SSIM, NCC, and LMSE, validating the framework's applicability to practical healthcare scenarios.
- The proposed method outperforms traditional spatial and transform domain steganography techniques in imperceptibility, payload capacity, and robustness to perturbations.

Recent works

Steganography, as a technique for embedding information within images, has evolved significantly in recent years to address increasing demands for data confidentiality in domains such as telemedicine and medical diagnostics. Classical spatial-domain methods like Least Significant Bit (LSB) substitution and Pixel Value Differencing (PVD) have been widely adopted due to their simplicity and low computational overhead. LSB-based approaches operate by modifying the least significant bits of pixel intensity values to conceal secret data, utilizing the human

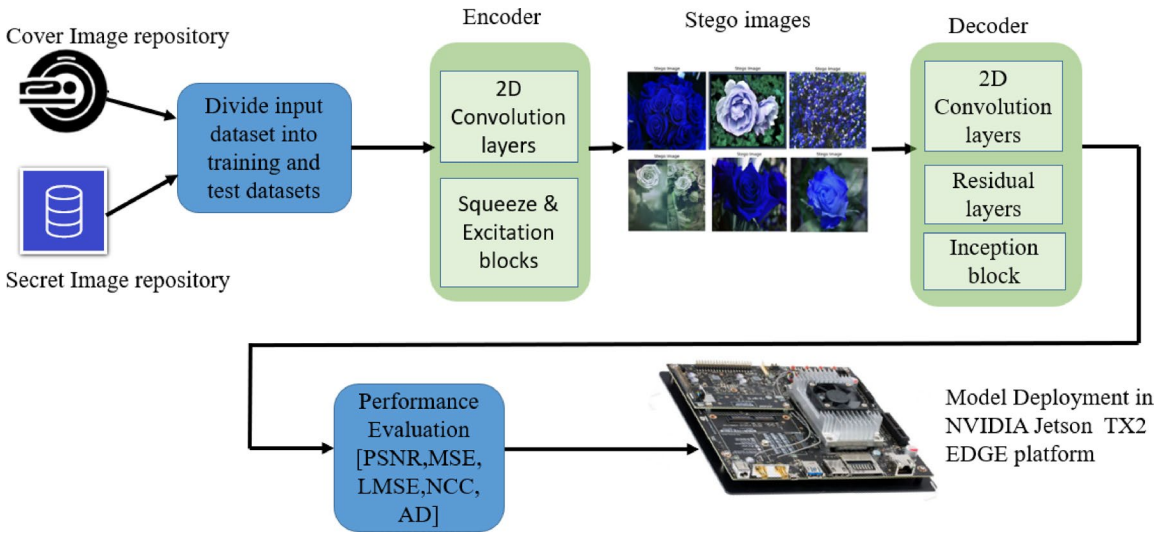

Fig. 1. Architecture of the proposed steganography approach.

visual system's insensitivity to small changes [5]. These methods provide minimal visual distortion and are useful in controlled environments. However, their vulnerability to statistical attacks, noise, format conversions, and compression artifacts restricts their use in real-world healthcare settings where robustness is essential [6,7].

To overcome these limitations, several enhancements to LSB have been proposed, such as LSB matching, adaptive embedding based on image intensity or edge features [6,7], and learning-based optimization strategies [8,9]. While some methods attempt to increase embedding capacity by expanding into multiple LSB planes [10], these often compromise imperceptibility which is an unacceptable trade-off in clinical imaging where even minor distortions may impact diagnosis.

Pixel Value Differencing (PVD) methods, in contrast, utilize differences between adjacent pixel values for embedding, offering improved imperceptibility by aligning with human visual sensitivity. Adaptive PVD schemes have further refined these techniques by predicting optimal embedding positions using various neighbourhoods models [11]. Yet, PVD remains vulnerable to histogram-based steganalysis and often lacks sufficient randomness in its embedding strategy. Security-enhanced variations such as improved rightmost digit replacement (iRMDR), Parity-Bit PVD (PBPVD) [12], and pseudo-random block selection [13] have been introduced to mitigate these issues. Hybrid models combining LSB and PVD [14,15] offer incremental improvements, but many still fall short of the robustness required for modern medical data transmission.

Transform-domain steganography techniques offer greater resilience against noise and signal processing operations. Methods such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), and Discrete Fourier Transform (DFT) embed data in the frequency components of images to improve imperceptibility and robustness [16]. DCT-based techniques embed data in mid-frequency components to balance robustness and visual fidelity, particularly in JPEG-compressed images [17]. But, conventional DCT approaches are prone to blocking artifacts and limited payload capacities. Adaptive DCT methods, which leverage coefficient variations across blocks, attempt to improve performance, but still face limitations in robustness and embedding efficiency.

DWT-based steganography decomposes images into multi-resolution sub-bands, embedding data in high-frequency components where alterations are less perceptible [18]. This makes DWT especially suitable for high-resolution modalities like CT and MRI, where structural fidelity is essential. Hybrid DWT-DCT methods combine the advantages of both transforms to further enhance security and efficiency. But they remain susceptible to geometric distortions and may introduce perceptual artifacts in sensitive regions of medical scans. DFT-based approaches manipulate phase and magnitude components for embedding. While phase modification is often imperceptible to the human eye, the global nature of the Fourier transform means that minor frequency changes can have broad spatial impact, making DFT less effective for localized or structure-preserving medical applications [19].

Recent advances in deep learning have opened new possibilities for adaptive, robust, and high-capacity steganographic systems. Convolutional neural networks (CNNs), attention mechanisms, and encoder-decoder frameworks enable content-aware embedding that maintains the diagnostic quality of medical images. These methods provide improved imperceptibility, robustness to distortions, and generalization across modalities such as CT, MRI, and ultrasound, making them well-suited for telemedicine and electronic health record (EHR) systems. Baluja [20] pioneered deep learning-based image steganography, demonstrating that a complete image can be hidden within another. In contrast, the resulting stego images often exhibited poor visual fidelity. Duan et al. [21] addressed this limitation using a U-Net architecture to extract and integrate hierarchical features, improving visual quality [22]. Yu et al. [23] introduced attention masks to identify optimal embedding regions, further enhancing imperceptibility.

Despite these advances, conventional CNNs often struggle with extracting fine-grained secret features, leading to distortions in reconstructed images. To mitigate this, Lu et al. [24] proposed the Invertible Steganography Network (ISN), employing parameter-sharing mechanisms for efficient embedding and extraction. Jing et al. [25] integrated a wavelet loss to guide embedding into high-frequency regions, preserving structural details. On the other hand, the lack of explicit attention mechanisms limited their ability to emphasize task-relevant features. Subsequent works, incorporated channel attention mechanisms to highlight salient features, while Li et al. [26] introduced spatial-channel joint attention within an invertible framework to minimize distortion. Still, many of these models rely on fixed-kernel convolutional attention, which limits their receptive field and adaptability. To overcome these constraints, we adopt Squeeze-and-Excitation (SE) blocks [27], which dynamically recalibrate channel-wise feature activations, enabling the model to focus on the most relevant information during both encoding and decoding. This improves the robustness of the embedded content while preserving the quality of both the stego and extracted medical images. Table 1 summarizes the strengths, weaknesses, and evaluation metrics of the major steganographic techniques discussed.

Methodology

This section describes the proposed deep learning-based steganography framework designed for secure, high-fidelity medical image embedding and reconstruction. Our model adopts a hybrid encoder–decoder architecture incorporating residual connections, Squeeze-and-Excitation (SE) blocks, dilated convolutions, and Inception modules to improve robustness, embedding accuracy, and imperceptibility. The entire system is optimized for real-time deployment on the NVIDIA Jetson TX2 edge computing platform. The notations used throughout this paper are summarized in Table 2.

Overall architecture

The proposed steganographic system comprises two core modules: an encoder network that embeds secret medical images into natural cover images, and a decoder network that reconstructs the hidden images with minimal perceptual distortion Both the encoder and decoder are built using deep convolutional operations

Technique	Strengths	Weaknesses	Evaluation Metrics
Classical (Spatial Domain) -LSB, PVD, LSB-Matching, Hybrid (LSB + PVD) [5–15]	-Low computational complexity -Easy to implement -High embedding capacity (esp. LSB) -Minimal visual distortion under ideal conditions	-Vulnerable to statistical attacks (e.g., RS, chi-square) -Poor robustness to compression, noise, and format conversion -Deterministic embedding strategy makes detection easier -Risk of introducing diagnostic artifacts in medical images	- MSE (Mean Squared Error) - Histogram Analysis
Transform Domain DCT, DWT, DFT, Hybrid (DWT + DCT) [16–19]	- High imperceptibility - Robust to compression, filtering, and noise—Compatible with JPEG formats - Frequency localization improves adaptive embedding - Effective for medical images (MRI, CT)	-Higher computational complexity - Limited payload capacity - Careful region selection needed to avoid diagnostic distortion - DFT has low spatial localization	- PSNR (Peak Signal-to-Noise Ratio) - SSIM (Structural Similarity Index) - NCC (Normalized Cross-Correlation) - Robustness to JPEG, noise, cropping
Deep Learning-Based CNNs, U-Net, ISN, Attention Models [20–26]	-Adaptive, content-aware embedding - High robustness to distortions (compression, noise)—Cross-modality applicability (MRI, CT, OCT) - Strong imperceptibility-payload trade-off - Hard to detect via traditional steganalysis	-Requires large training datasets - High GPU and memory requirements - Potential for colour/texture distortion - Training may be unstable - Interpretability can be limited	- PSNR (Peak Signal-to-Noise Ratio) - SSIM (Structural Similarity Index) - Robustness to compression, noise, geometric distortions

Table 1. Comparison of Steganographic Techniques in Medical Imaging.

Notation	Description
I_c	Cover Image used for embedding
I_s	Secret medical image to be embedded
I_{concat}	Concatenated Tensor
$W_1, W_2, \ldots$	Weight matrices
$B_1, B_2, \ldots$	Bias
$*$	Convolution operation
$\delta(\cdot)$	ReLU Activation
$\sigma(\cdot)$	Sigmoid function
$\text{Conv}_{k\times k}^{(dilation=r)}$	Dilated convolution with kernel size $k \times k$ and rate r
F	Intermediate feature map
$F_1, F_2, \ldots$	Feature maps at various stages of the encoder or decoder
F_r	Residual feature map
F_{agg}	Aggregated feature map from skip connections
$F_{\text{dilated}}^{(r)}$	Output of dilated convolution with dilation rate r
z	Channel-wise descriptor after global average pooling
s	Channel-wise scaling factors from SE block
$\widehat{F}_{i,j,c}$	Recalibrated feature map after applying SE attention
$F_{1\times1}, F_{3\times3}, F_{5\times5}$	Feature maps from Inception module with different kernel sizes
F_{concat}	Concatenated multi-scale feature map from Inception module
I_{encoded}	Output stego image after encoding module
I_{decoded}	Reconstructed secret image from the decoder

Table 2. Summary of notations used in the proposed model.

augmented with attention and multi-scale feature learning techniques. Figure 2 shows the block diagram of the overall architecture and we give a detailed description of each module in subsequent sections.

Encoder architecture

The encoder receives a pair of inputs: a cover image I_c and a secret medical image I_s, each resized to the same spatial dimensions. The two inputs are then concatenated along the channel axis to form a unified tensor:

$$I_{concat} = \text{Concat}(I_c, I_s) \tag{1}$$

This tensor is passed through a sequence of convolutional layers initialized with 3×3 kernels and 64 filters. These layers extract low-level features from the combined representation. Each convolution operation is followed

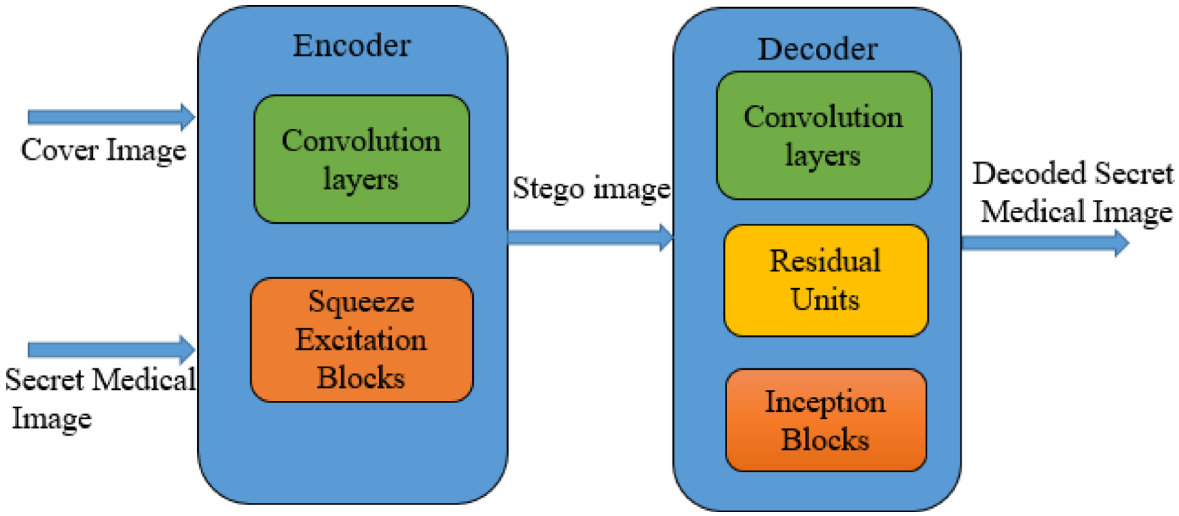

Fig. 2. Block diagram of the proposed encoder-decoder architecture.

by ReLU activation, batch normalization, and, in some blocks, residual skip connections to preserve gradient flow. Figure 3 shows the different layers in the encoder network.

Squeeze-and-excitation attention mechanism: To recalibrate channel-wise feature responses, SE blocks are integrated within the encoder. Given an intermediate feature map $F \in \mathbb{R}^{H\times W\times C}$, SE attention performs the following steps:

Squeeze: Apply global average pooling to obtain a channel descriptor:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j,c} \tag{2}$$

Excitation: Pass through two fully connected layers with non-linear activation:

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \tag{3}$$

Scaling: Rescale the original feature map:

$$\widehat{F}_{i,j,c} = s_c \cdot F_{i,j,c} \tag{4}$$

This attention module enhances the model's ability to focus on semantically meaningful features relevant for embedding. Figure 4 shows the different elements in the Squeeze and Excitation module.

Dilated convolutions: To capture multi-scale spatial information without significantly increasing computational cost, dilated convolutions are employed with rates $r = \{2, 4, 8\}$. These increase the receptive field and enable the model to encode contextual features at varying levels:

$$F_{dilated}^{(r)} = \text{Conv}_{3\times 3}^{(dilation=r)}(F) \tag{5}$$

Smaller rates capture texture and edges, while larger ones encode global patterns. Outputs from different dilation levels are combined using summation with residual projections.

Skip connections and feature aggregation: To facilitate deeper learning and enhance training stability, residual skip connections are introduced between intermediate feature maps. Intermediate outputs F_1, F_2 are aligned using 1×1 convolutions and aggregated as:

$$F_{agg} = \text{Conv}_{1\times 1}(F_1) + \text{Conv}_{1\times 1}(F_2) \tag{6}$$

This aggregation improves feature fusion and gradient propagation, preserving both local and global features critical to accurate embedding.

Encoded output generation: The aggregated features are passed through a final convolutional layer followed by a sigmoid activation to produce the stego image $I_{encoded}$:

$$I_{encoded} = \sigma(\text{Conv}_{3\times 3}(F_{agg})) \tag{7}$$

The stego image retains the same shape as the cover image and visually resembles it while imperceptibly containing the embedded medical content.

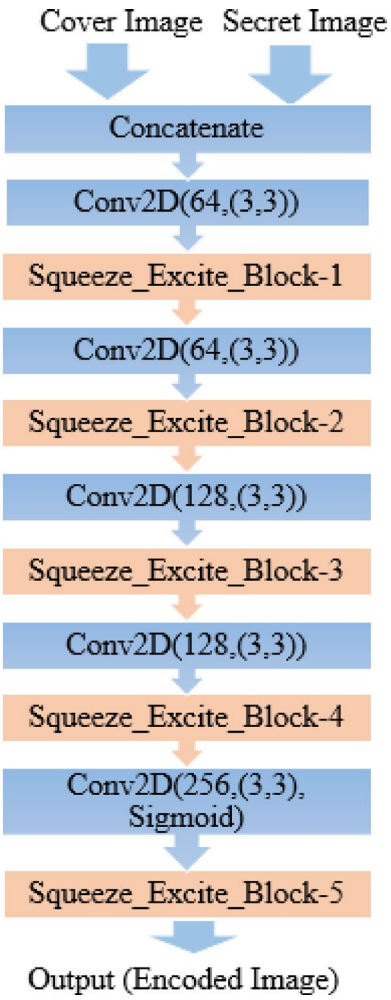

Fig. 3. Encoder Architecture.

Decoder architecture

The decoder module reconstructs the hidden medical image from the encoded stego image while preserving its structural and visual integrity. This ensures that the embedded diagnostic data is accurately recovered, with minimal distortion or loss of information. Figure 5 shows the different layers in the decoder architecture.

Initial feature extraction: The decoding process begins by passing the encoded image I_{stego} into a convolutional layer with a 3×3 kernel and 64 filters. This operation captures low-level spatial patterns required for subsequent reconstruction. The ReLU activation function introduces non-linearity, enhancing the feature representation. The transformation is mathematically expressed as:

$$F_1 = \mathrm{ReLU}(W_1 * I_{stego} + B_1) \tag{8}$$

Residual feature mapping: To improve gradient flow and learning stability, residual blocks are incorporated into the decoder. Each residual unit contains two convolutional layers with ReLU activations and a shortcut connection. This structure enables identity mappings that help preserve fine-grained details during decoding. The residual transformation can be described as follows:

$$F_{r1} = \mathrm{ReLU}(W_{r1} * F_{in} + B_{r1}) \tag{9}$$

$$F_{r2} = W_{r2} * F_{r1} + B_{r2} \tag{10}$$

If the input and output dimensions differ, a 1×1 convolution is applied to align the channel dimensions before the residual summation:

$$F_{out} = F_{r2} + \mathrm{Conv1x1}(F_{in}) \tag{11}$$

This residual structure facilitates effective feature learning and reconstruction of complex anatomical structures from the stego image. Figure 6 shows the architecture of the residual block.

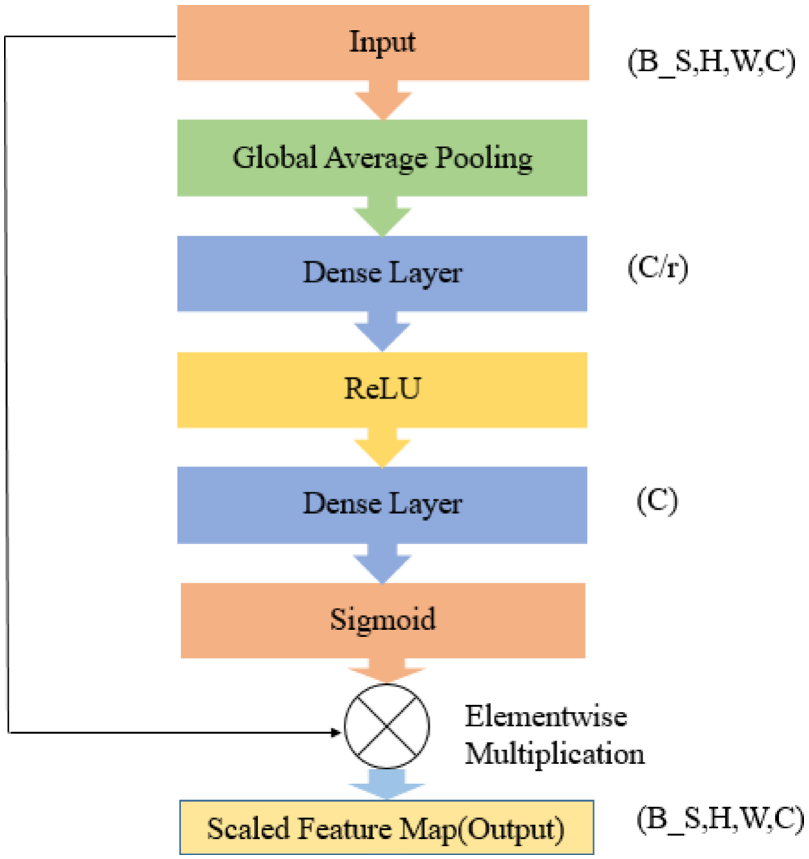

Fig. 4. Squeeze and Excitation Attention mechanism.

Multi-scale decoding via inception modules: Following residual mapping, the decoder employs Inception modules to extract multi-scale spatial features. These blocks apply convolutional filters of varying kernel sizes ($1 \times 1, 3 \times 3$, and 5×5) in parallel, enabling the model to capture both fine and coarse details within the image which is shown in Fig. 7. The individual operations are defined as:

$$F_{1x1} = \text{ReLU}(W_{1x1} * F + B_{1x1}) \tag{12}$$

$$F_{3x3} = \text{ReLU}(W_{3x3} * F + B_{3x3}) \tag{13}$$

$$F_{5x5} = \text{ReLU}(W_{5x5} * F + B_{5x5}) \tag{14}$$

These feature maps are concatenated along the channel dimension to form the final multi-scale representation:

$$F_{concat} = [F_{1x1}, F_{3x3}, F_{5x5}] \tag{15}$$

The use of inception structures ensures that relevant spatial patterns at multiple resolutions are incorporated into the reconstruction process, improving robustness to variations in medical image structure.

Final reconstruction and output: A final 3×3 convolutional layer, followed by a sigmoid activation function, is applied to generate the output secret image. This ensures that the reconstructed pixel values remain within the normalized range $[0, 1]$, suitable for grayscale medical imaging:

$$I_{decoded} = \sigma(W_f * F_{concat} + B_f) \tag{16}$$

Decoder enables precise and high-fidelity reconstruction of the embedded medical image, making it suitable for telemedicine and real-time diagnostic applications on edge devices.

Experiments and results

This section presents the experimental evaluation of the proposed deep learning-based steganography model, focusing on reconstruction fidelity, robustness, and computational efficiency. A series of experiments were conducted to validate the performance of the proposed architecture on diverse datasets and under varying perturbation scenarios. Metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Mean Squared Error (MSE), and Normalized Cross-Correlation (NCC) were employed to quantify performance.

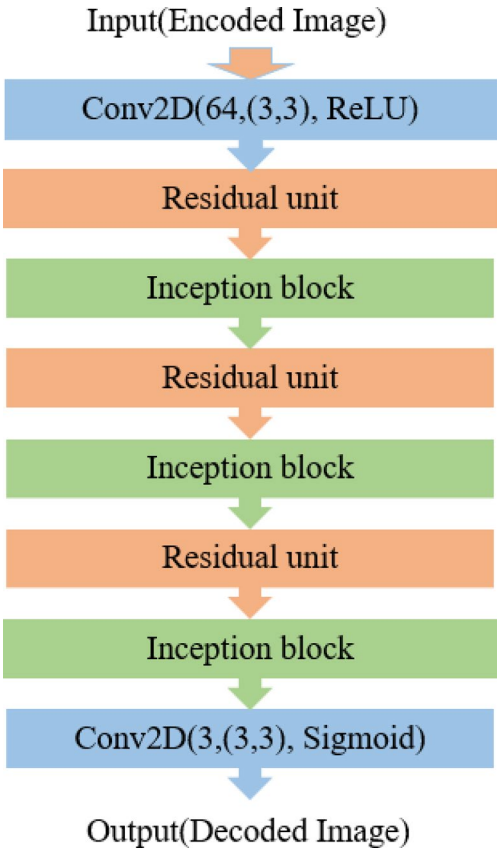

Fig. 5. Decoder Architecture.

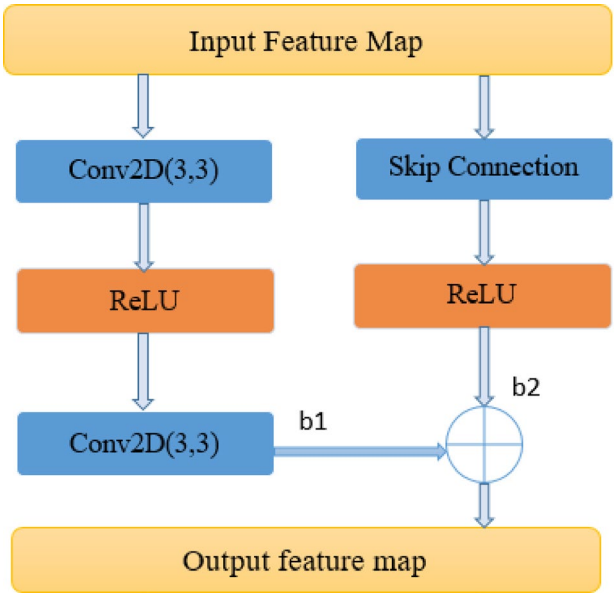

Fig. 6. Architecture of a Residual block.

Hardware and software configuration

All experiments were implemented in Python 3.8 using the PyTorch framework. The model was trained and tested on a workstation equipped with an NVIDIA Jetson TX2 platform, a power-efficient embedded system featuring a 256-core Pascal GPU and a quad-core ARM CPU. CUDA and cuDNN were used to enable hardware

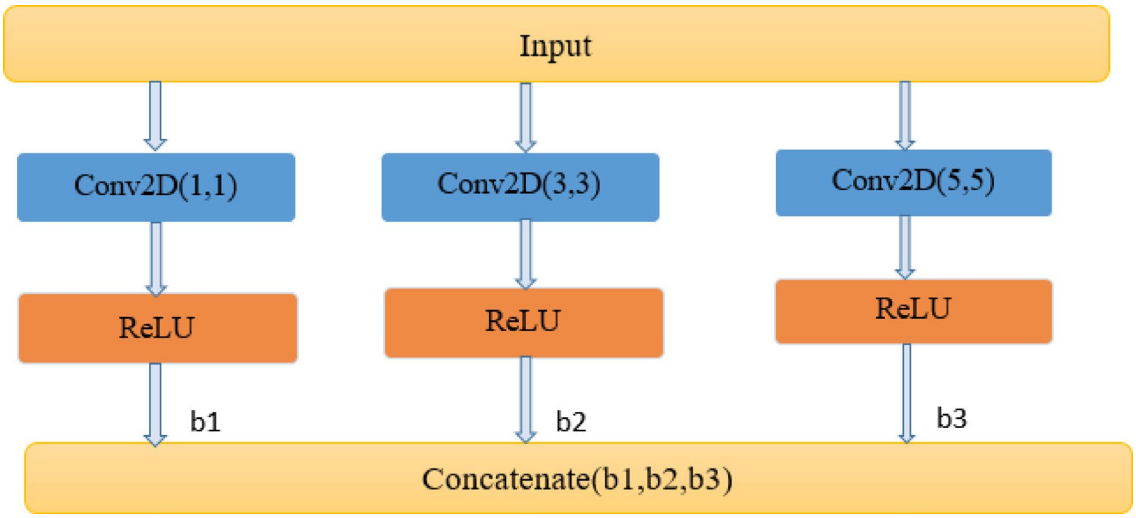

Fig. 7. Architecture of an Inception module.

acceleration. Hyperparameter tuning and training monitoring were conducted using PyTorch Lightning and TensorBoard.

Datasets

To evaluate the performance of the proposed model, three distinct datasets were utilized. The cover images comprised 600 natural flower images [28,29], serving as the visual medium for embedding. Two types of medical datasets were employed as secret images: 600 MRI brain scans used for tumour detection [30], and 600 Optical Coherence Tomography (OCT) eye images used for glaucoma detection [31].

All images were pre-processed to a standardized resolution of $256 \times 256 \times 3$ pixels to ensure uniformity in input dimensions and compatibility with the encoder-decoder framework. The datasets were partitioned into training and testing subsets using an 80:20 split. The network was trained for 200 epochs using a batch size of 8. Optimization was performed using the Adam algorithm, with an initial learning rate set to 0.0001. Figure 8 shows sample images- cover image, secret MRI image, stego image and decoded secret image corresponding to three samples from MRI brain dataset and Fig. 9 shows three samples from OCT Glaucoma dataset.

Quantitative evaluation

The effectiveness of the model was evaluated using a range of standard metrics including PSNR, MSE, Structural Content (SC), Normalized Cross-Correlation (NCC), Maximum Difference (MD), Laplacian Mean Squared Error (LMSE), Normalized Absolute Error (NAE), and Structural Similarity Index (SSIM). These metrics were computed for all test images and box plots were plotted across 25 randomly selected test images from the brain MRI dataset to assess distortion, similarity, and structural preservation.

Peak signal-to-noise ratio (PSNR): PSNR is a widely used metric that quantifies the reconstruction fidelity by measuring the pixel-level distortion introduced during embedding.

$$\mathrm{PSNR} = 10 \cdot \log_{10} \left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}} \right) \tag{17}$$

where MAX is the maximum pixel value (255 for 8-bit images), and MSE denotes the Mean Squared Error between the secret and reconstructed images.

Our model achieved average PSNR values of 39.02 dB (brain MRI) and 38.75 dB (glaucoma), indicating high-fidelity reconstruction. Figure 10a shows the box plot of PSNR values obtained for MRI Brain image dataset for 25 test images. The median is centered around 39 dB, with most values falling in the 38–40 dB range. A single upper outlier above 42 dB is observed, suggesting occasional enhanced reconstruction due to variations in embedding strength.

Mean squared error (MSE): MSE measures the average squared difference between the original and decoded pixel intensities.

$$\mathrm{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(I_s\left(i,j\right) - I_{decoded}\left(i,j\right) \right)^2 \tag{18}$$

Lower MSE values indicate minimal distortion. Figure 10b shows a median value of ~ 9, with most values ranging from 7 to 12 obtained for MRI Brain image dataset for 25 test images. The absence of extreme outliers reflects the stability of the model.

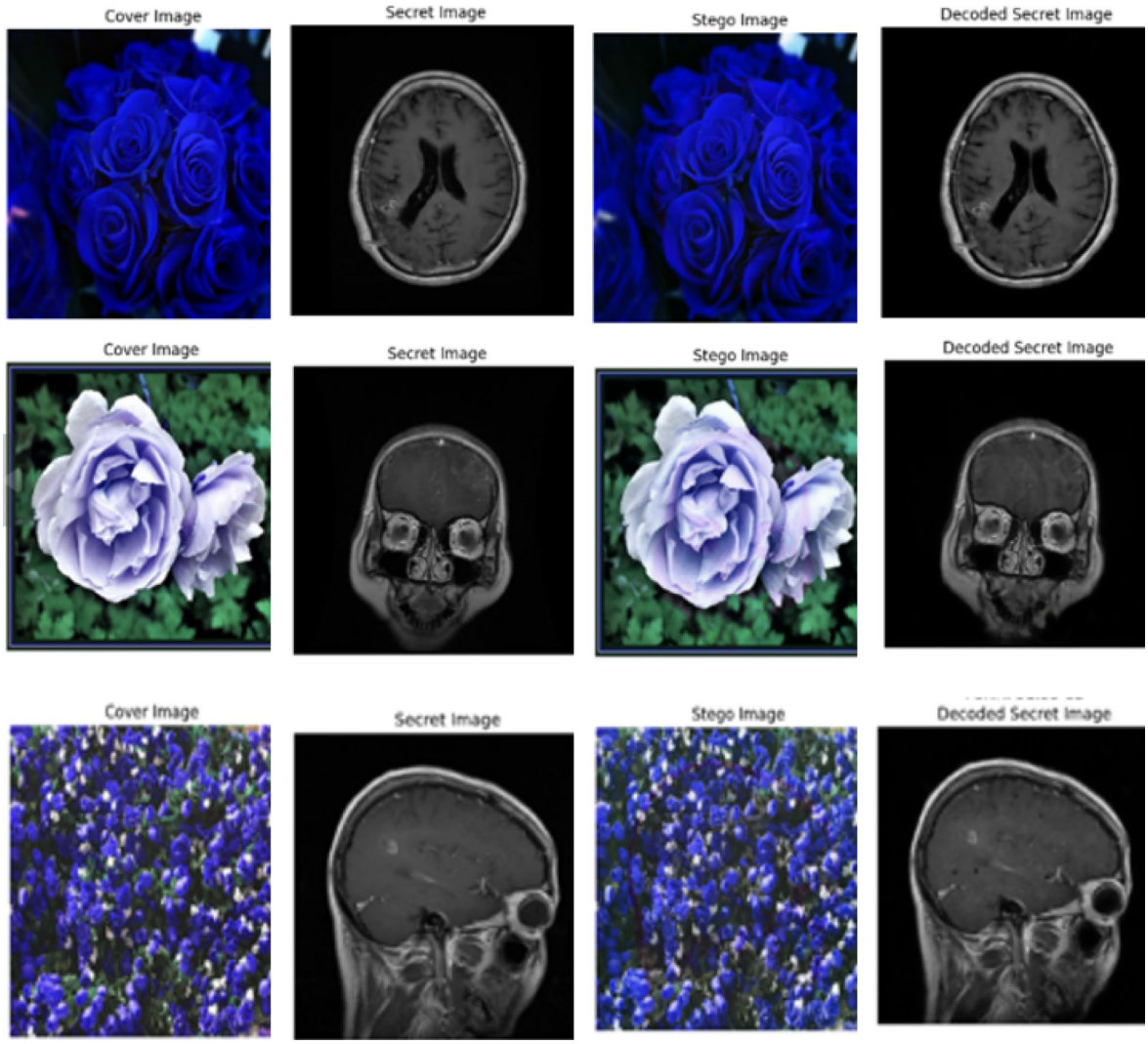

Fig. 8. Cover image, secret image, stego image & decoded secret image of three MRI Brain images.

Structural content (SC): SC assesses structural preservation by comparing the energy of the original and decoded images.

$$\mathrm{SC} = \frac{\sum I_s(i,j)^2}{\sum I_{decoded}(i,j)^2} \tag{19}$$

Values close to 1 indicate minimal deviation. As seen in Fig. 10c, the majority of SC values are centered around 1.05, indicating excellent preservation of structural content, which is very important for medical imaging.

Normalized cross-correlation (NCC): NCC measures pixel-level similarity between the secret and decoded images:

$$\mathrm{NCC} = \frac{\sum I_s(i,j) \cdot I_{decoded}(i,j)}{\sum I_s(i,j)^2} \tag{20}$$

Ideal values are close to 1. Figure 11a illustrates NCC values predominantly above 0.96, with a median near 0.97 and a single upper outlier at 1.0, confirming excellent reconstruction fidelity.

Maximum difference (MD): MD quantifies the largest absolute pixel-wise difference:

$$\mathrm{MD} = \max_{i,j} |I_s(i,j) - I_{decoded}(i,j)| \tag{21}$$

Lower values are preferred for medical images. As shown in Fig. 11b, most MD values lie between 25 and 35, with a median around 30. A few samples exhibit higher deviation, likely due to localized image complexity.

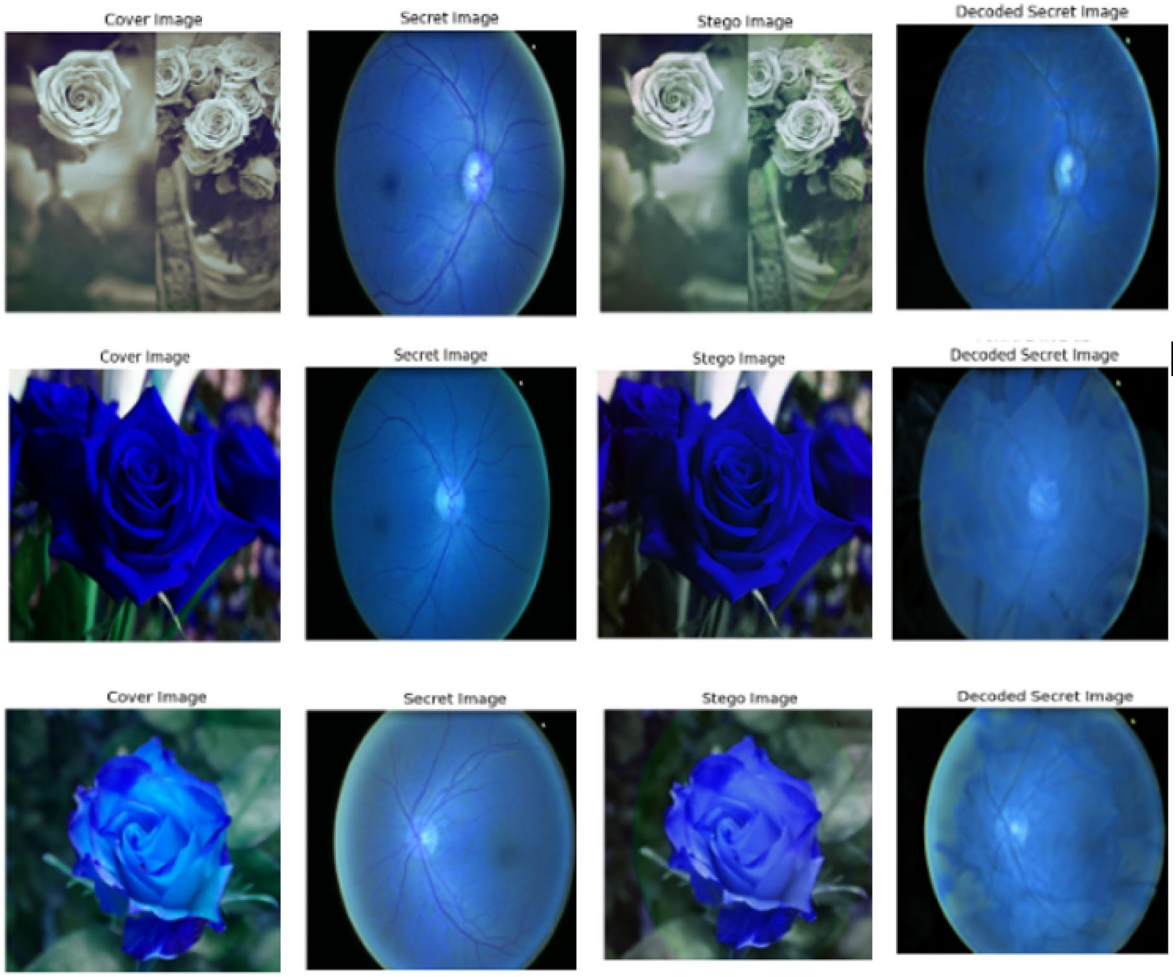

Fig. 9. Cover image, secret image, stego image & decoded secret image of three OCT Glaucoma images.

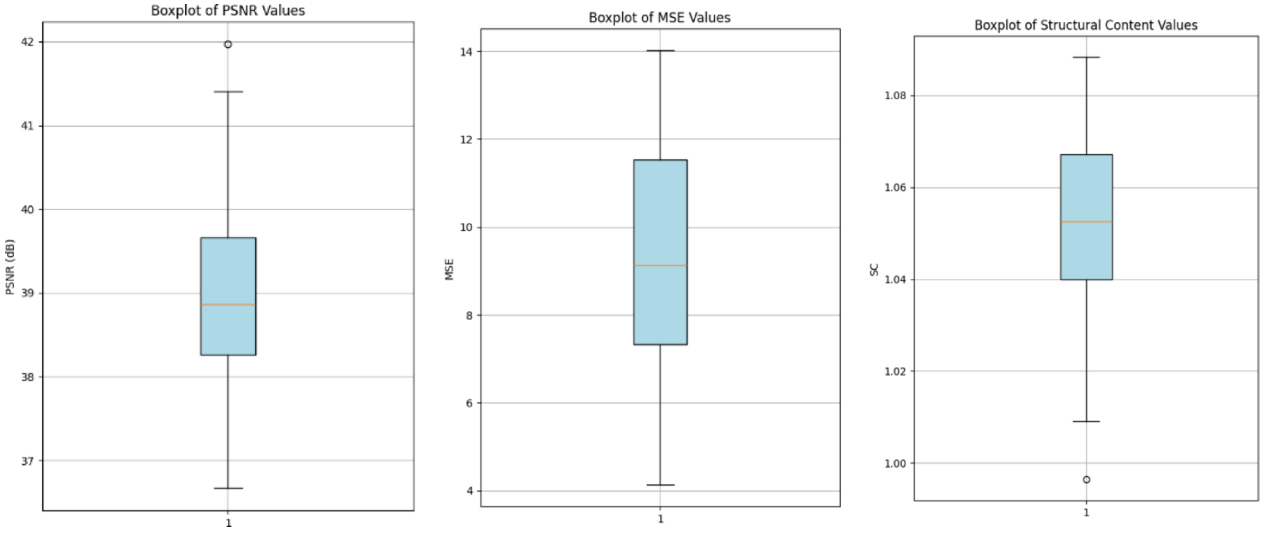

Fig. 10. Box plots of (**a**) PSNR (**b**) MSE and (**c**) SC for 25 MRI brain test images.

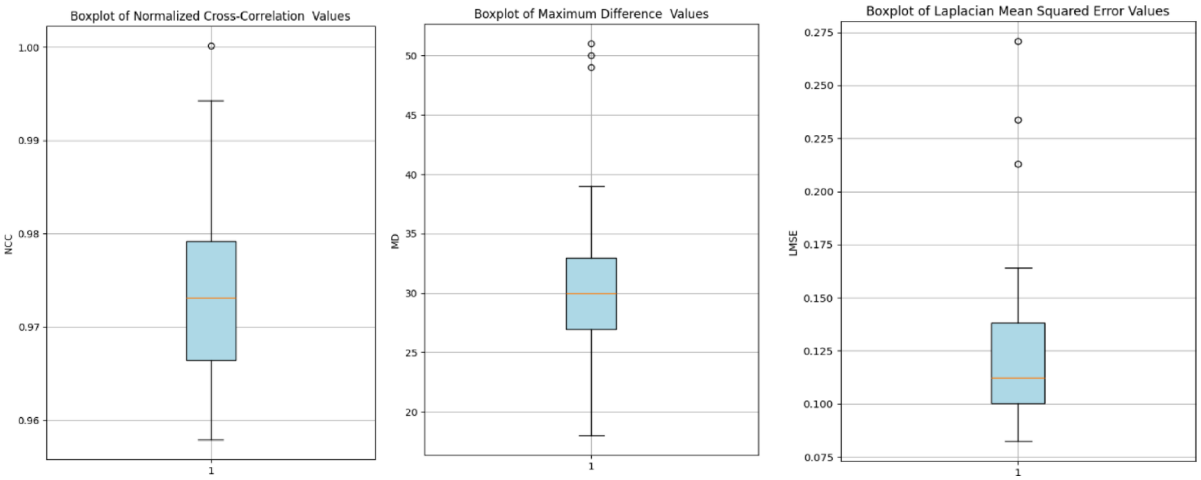

Fig. 11. Box plots of (**a**) NCC (**b**) MD and (**c**) LMSE for 25 MRI brain test images.

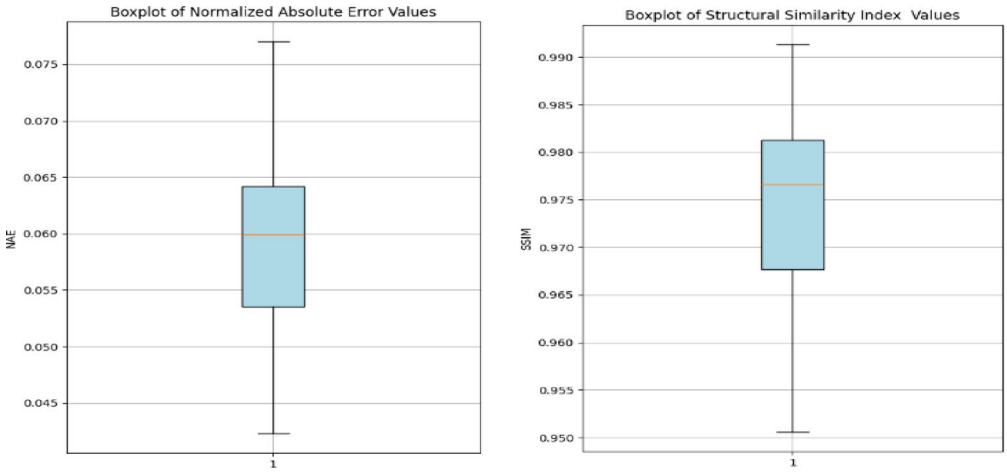

Fig. 12. Box plot on brain MRI images (**a**) NAE values (**b**) SSIM values for 25 MRI brain test images.

Laplacian mean squared error (LMSE): LMSE evaluates edge-based distortion using the Laplacian of the image:

$$\mathrm{LMSE} = \frac{\sum \left[\nabla^2 I_s(i,j) - \nabla^2 I_{decoded}(i,j)\right]^2}{\sum \left[\nabla^2 I_s(i,j)\right]^2} \tag{22}$$

A low LMSE (< 0.1) signifies excellent structural preservation. Figure 11c shows LMSE values predominantly in the 0.09–0.15 range, with a median near 0.11.

Normalized absolute error (NAE): NAE measures perceptual distortion between the secret and decoded image.

$$\mathrm{NAE} = \frac{\sum |I_s(i,j) - I_{decoded}(i,j)|}{\sum |I_s(i,j)|} \tag{23}$$

NAE values < 0.05 are considered excellent. Figure 12a shows a median NAE of ~ 0.06 with low variability, indicating good imperceptibility across test cases.

Structural similarity index measure (SSIM): SSIM evaluates perceptual similarity based on luminance, contrast, and structure.

$$\mathrm{SSIM}(I_s, I_d) = \frac{(2\mu_s\mu_{decoded} + C_1)(2\sigma_{s,decoded} + C_2)}{(\mu_s^2 + \mu_{decoded}^2 + C_1)(\sigma_s^2 + \sigma_{decoded}^2 + C_2)} \tag{24}$$

with μ, σ, and σ_{sd} representing means, variances, and covariances, respectively.

As shown in Fig. 12b, the SSIM values for the proposed model are consistently high, with a median near 0.98 and a minimum above 0.95. These results confirm minimal perceptual distortion and excellent structural consistency.

Robustness analysis under common image perturbations

To evaluate the robustness of the proposed method against common image distortions, we applied four types of perturbations- JPEG compression, Gaussian noise, cropping with resizing, and salt & pepper noise. The quantitative results are summarized in Fig. 13 which presents a heatmap comparing PSNR and SSIM across these perturbations. It illustrates that the model maintains high perceptual similarity and signal fidelity even under moderate perturbations.

- JPEG Compression (Q = 75) yields the highest PSNR (39.02 dB) and SSIM (0.975), indicating minimal perceptual degradation which suggests that our technique is highly resilient to lossy compression artifacts.
- Gaussian Noise ($\sigma = 0.01$) moderately impacts performance, with a PSNR of 37.55 dB and SSIM of 0.963, reflecting slight pixel-level perturbations but relatively preserved structural integrity.
- Cropping + Resizing resulted in a more noticeable degradation (PSNR = 36.10 dB, SSIM = 0.950), likely due to interpolation and information loss at borders.
- Salt & Pepper Noise imposes the greatest challenge, with the lowest PSNR (34.70 dB) and SSIM (0.931), attributable to the high-intensity, localized nature of the noise.

So, our model is robust especially under realistic distortions like compression and Gaussian noise. But the performance slightly degrades under spatial transformations and impulsive noise, warranting future improvements in noise-aware training schemes.

Ablation study

To evaluate the individual contributions of key architectural components within the proposed steganographic model, an ablation study was performed. In each experiment, one or more components were removed or modified while keeping the remaining architecture unchanged. The effect of each modification was assessed based on the average PSNR and SSIM values over the test set, allowing for a quantitative analysis of image fidelity and structural preservation.

Model A: removal of residual connections—Residual connections were excluded from the encoder-decoder architecture to assess their influence on feature propagation. The absence of skip connections resulted in a PSNR of 34.12 dB and SSIM of 0.915. The degradation in performance highlights the critical role of residual learning in maintaining deep feature retention and facilitating gradient flow, which are essential for accurate reconstruction.

Model B: exclusion of squeeze-and-excitation (SE) blocks—SE blocks were removed while preserving all other architectural elements. The modified model yielded a PSNR of 35.05 dB and SSIM of 0.927. This decline demonstrates the importance of channel-wise attention in recalibrating feature responses, allowing the network to selectively emphasize salient features during both embedding and decoding.

Model C: replacement of dilated convolutions with standard convolutions—In this variant, all dilated convolutions were substituted with standard convolutions. A significant drop in performance was observed, with PSNR falling to 33.80 dB and SSIM to 0.910. These results suggest that dilated convolutions are instrumental

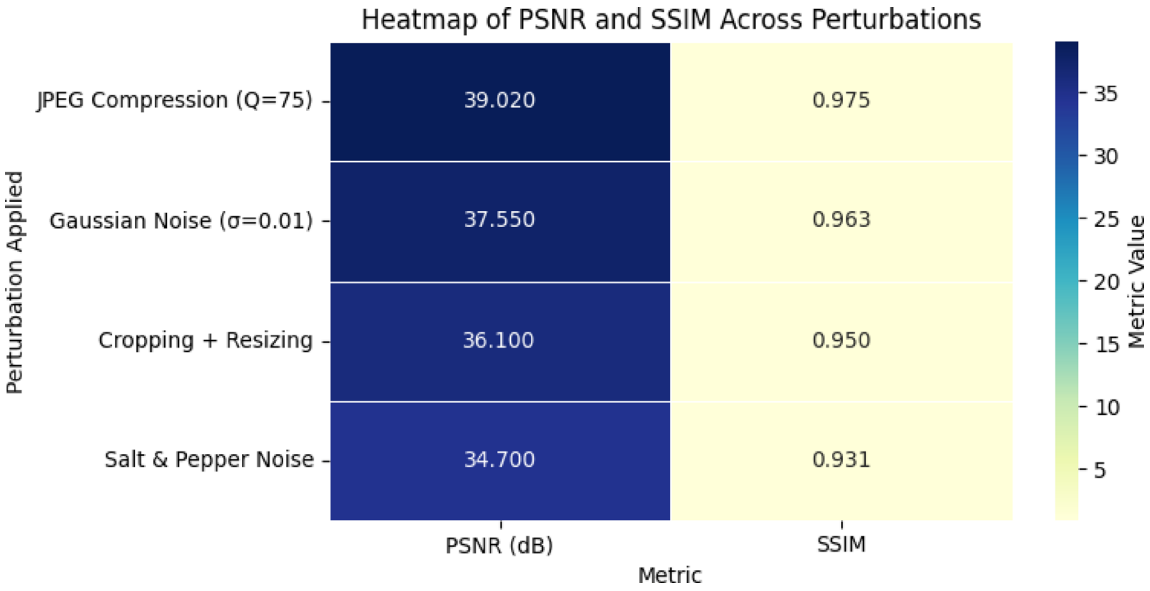

Fig. 13. Heatmap of PSNR (dB) and SSIM values across four perturbation types. Darker blue indicates higher PSNR, while lighter shades in SSIM indicate higher structural similarity.

in expanding the receptive field without increasing the parameter count, thereby enhancing spatial feature extraction necessary for effective steganographic encoding.

Model D: removal of inception modules—The inception blocks responsible for multi-scale feature extraction were eliminated. The resulting model achieved a PSNR of 34.50 dB and SSIM of 0.918. This indicates that the parallel use of multiple kernel sizes contributes substantially to the model's adaptability in capturing hierarchical spatial patterns, which supports better feature representation across varied image textures.

Model E: Joint removal of SE blocks and dilated convolutions—To assess the combined effect of removing both channel attention and spatial context mechanisms, SE blocks and dilated convolutions were jointly removed. The model's performance further deteriorated, with PSNR at 33.30 dB and SSIM at 0.905. The result underscores the complementary role these components play in balancing local emphasis and global context, both crucial for accurate and imperceptible data embedding.

Model F: elimination of all attention mechanisms—All attention mechanisms, including SE blocks and channel-based recalibration layers, were removed in this configuration. The performance declined to a PSNR of 32.90 dB and SSIM of 0.895. These findings emphasize the role of attention in enhancing discriminative features, improving the network's capability to recover fine details from stego images.

Model G: removal of residual and attention mechanisms—In the most constrained configuration, both residual connections and all attention mechanisms were removed. This led to the most pronounced performance degradation, with a PSNR of 32.20 dB and SSIM of 0.880. The simultaneous exclusion of deep feature retention and selective attention mechanisms severely impacted the model's ability to reconstruct the hidden content, reaffirming their synergistic importance in achieving high-fidelity steganography.

Hardware implementation and deployment feasibility

The proposed steganography model was deployed on the NVIDIA Jetson TX2 platform to evaluate its real-time processing capability, energy efficiency, and suitability for edge applications in telemedicine and medical imaging. The Jetson TX2 is equipped with a 256-core Pascal GPU, a 6-core ARM CPU, and 8 GB of LPDDR4 RAM, making it a compact yet powerful embedded system ideal for low-power, real-time AI tasks.

The architecture's integration of residual connections, squeeze-and-excitation (SE) blocks, inception modules, and dilated convolutions enables high-capacity embedding while maintaining visual fidelity. Residual learning facilitates deep feature propagation, SE blocks enhance attention-based refinement, and inception modules allow multi-scale spatial feature extraction all of which are essential for robust and accurate reconstruction of medical images.

During inference, convolutional operations, adaptive embedding, and decoding tasks were significantly accelerated by the Jetson TX2's parallel CUDA cores. The system consistently achieved a processing time of 25–35 ms per image and a throughput of approximately 25–30 images per second. These results demonstrate the feasibility of deploying the model for real-time steganographic operations in mobile or low-resource medical environments. Table 3 summarizes the computational resource usage and environmental suitability of the proposed model on the Jetson TX2 platform.

Benchmarking with State-of-the-Art Techniques

To evaluate the performance of the proposed steganographic model, we conducted a comparative analysis with recent state-of-the-art methods, including HiDDeN [32], SteganoGAN [33], HCISNet [34], and AVGGAN [35]. The comparison was performed using three widely adopted image quality metrics- PSNR, MSE, and SSIM-on both the brain MRI and OCT glaucoma datasets. Table 4 shows the corresponding metric values obtained for three sample images from MRI Brain image datset.

The proposed model consistently achieved the highest or near-highest PSNR values, ranging from 38.51 dB to 41.96 dB, indicating excellent reconstruction fidelity. SSIM values remained above 0.96 for all test cases, with

Sl No	Category	Details	Feasibility
1	Hardware	NVIDIA Jetson TX2 (256-core Pascal GPU, 6-core ARM CPU, 8 GB LPDDR4 RAM)	Compact and powerful edge AI hardware, suitable for medical use
2	Power Consumption	7.5–15 W (under moderate-to-high inference load)	Low enough for mobile clinics or battery-powered diagnostic tools
3	Energy Efficiency	~0.5 W/image (based on 15W max power and 30 images/sec throughput)	Energy-efficient for real-time steganography in portable and embedded setups
4	Model Size	~25–30 MB	Lightweight for TX2's onboard storage
5	Disk Usage	~3.0 GB (model, dataset, dependencies)	Easily accommodated on TX2's internal or external storage
6	RAM Usage	~1.8–2.2 GB during inference	Fits comfortably within 8 GB RAM
7	Processing Time (Latency)	~25–35 ms per image (encoder + decoder on Jetson TX2 with CUDA acceleration)	Real-time embedding and decoding achievable
8	Processing Throughput	~25–30 images/sec	Suitable for continuous or batch secure image transmission
9	Environmental Constraints	0–50 °C	Suitable for mobile labs, rural health units, or static clinical environments
10	Deployment feasibility	Plug-and-play deployment	High-no external dependencies beyond standard CUDA stack

Table 3. Computational Resource Consumption and Real-World Feasibility of the Proposed Model on NVIDIA Jetson TX2.

Dataset	Metric	Image	HiDDeN [32]	SteganoGAN [33]	HCISNet [34]	AVG-GAN [35]	Proposed
MRI Brain Dataset	PSNR	MRI-1	36.96	36.56	38.86	39.95	40.3
		MRI-2	37.25	35.96	37.65	39.97	41.09
		MRI-3	36.85	36.02	38.99	40.25	41.96
	MSE	MRI-1	6.28	6.75	4.04	2.57	6.05
		MRI-2	5.91	7.55	5.42	2.54	5.05
		MRI-3	6.4	7.62	3.91	2.61	4.13
	SSIM	MRI-1	0.96	0.96	0.97	0.98	0.98
		MRI-2	0.97	0.96	0.97	0.98	0.99
		MRI-3	0.96	0.96	0.97	0.98	0.98
OCT Glaucoma Dataset	PSNR	OCT-1	36.86	36.56	38.16	38.95	38.51
		OCT-2	37.15	36.22	37.44	38.47	38.85
		OCT-3	38.45	37.23	38.44	38.13	39.71
	MSE	OCT-1	6.34	6.75	4.85	4.09	9.14
		OCT-2	5.79	7.54	5.17	4.26	8.46
		OCT-3	4.27	5.93	4.28	3.39	6.94
	SSIM	OCT-1	0.96	0.96	0.97	0.98	0.98
		OCT-2	0.96	0.95	0.97	0.98	0.96
		OCT-3	0.97	0.96	0.97	0.98	0.98

Table 4. Comparison of Key metrics with State-of-the-Art Techniques.

Reference	SSIM (MRI)	PSNR (MRI)	SSIM (OCT)	PSNR (OCT)
HiDDeN	0.96	36.22	0.96	35.13
SteganoGAN	0.84	36.46	0.83	36.02
HCISNet	0.92	38.87	0.91	37.55
AVG-GAN	0.98	39.58	0.975	38.18
Proposed	0.975	39.02	0.975	38.75

Table 5. Average SSIM and PSNR values over 25 test images for brain MRI and glaucoma OCT datasets.

peaks of 0.99 for brain MRI samples, highlighting strong preservation of structural information. While MSE values were slightly higher in some OCT cases compared to AVGGAN, the overall visual quality remained high, as reflected by superior SSIM and PSNR values.

To further assess model generalizability and average performance across a broader test set, we computed the mean PSNR and SSIM values over 25 randomly selected images from each dataset. As shown in Table 5, the proposed method consistently outperformed SteganoGAN, HCISNet, and CSIS on both brain MRI and glaucoma OCT datasets. Its performance closely matched AVGGAN, achieving competitive SSIM and PSNR values while maintaining lower model complexity and real-time deployability.

Conclusion and future works

In this research, we introduced a novel deep learning-based steganographic framework designed specifically for secure medical image transmission in telemedicine and teleradiology applications. By using Squeeze-and-Excitation (SE) blocks, Inception modules, and residual connections within an encoder-decoder architecture, the proposed model effectively balances embedding capacity, imperceptibility, and reconstruction accuracy. Experimental results on MRI brain scans and OCT glaucoma datasets demonstrated high visual fidelity of the stego images and successful reconstruction of secret image, with average PSNR values exceeding 39 dB and SSIM values approaching 0.98. Unlike traditional steganographic techniques, our model is robust against various perturbations and is optimized for real-time edge deployment on NVIDIA Jetson TX2, addressing the practical constraints of clinical environments. The integration of dilated convolutions and multi-scale feature extraction further enhances the system's ability to capture spatial context and maintain diagnostic integrity, making it suitable for high-resolution medical imagery.

This work establishes a viable pathway for implementing secure, lightweight, and scalable steganographic solutions in resource-constrained settings, where both data confidentiality and real-time processing are critical. Future research may explore integrating transformer-based attention mechanisms and cross-modality steganography to further increase adaptability across diverse medical imaging standards.

Data availability

The datasets generated and/or analysed during the current study are available in the following repositories. Cover image datasets- https://www.kaggle.com/datasets/imsparsh/flowers-dataset and https://www.kaggle.com/dat

asets/muhammedtausif/rose-flowers. Secret image dataset of MRI brain images-https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection. Secret image dataset of OCT scan images of eye for Glaucoma detection-: https://www.kaggle.com/datasets/sshikamaru/glaucoma-detection

Received: 10 April 2025; Accepted: 14 August 2025
Published online: 25 August 2025

References

1. Kumar, S. A survey of machine learning applications in medical imaging for neurodegenerative disease diagnosis. *IET Conf. Proc.* **2024**(23), 274–281 (2024).
2. Akter, Mst Shamima, et al. "Big Data Analytics in Healthcare: Tools, Techniques, And Applications-A Systematic Review." *Journal of Next-Gen Engineering Systems* 2.01 (2025): 29–47.
3. Oluseyi, Omotosho Moses, Akpan Itoro Udofot, and Edim Bassey. "Enhancing Medical Image Diagnosis Using Convolutional Neural Network and Transfer Learning."
4. Pandey, Aparna, et al. "Innovations in AI and ML for Medical Imaging: An Extensive Study with an Emphasis on Face Spoofing Detection and Snooping." *The Impact of Algorithmic Technologies on Healthcare* (2025): 127–155.
5. Kamil Khudhair, S.; Sahu, M.; K. R., R.; Sahu, A.K. Secure Reversible Data Hiding Using Block-Wise Histogram Shifting. *Electronics* **2023**, *12*, 1222. https://doi.org/10.3390/electronics12051222
6. Luo, W., Huang, F. & Huang, J. Edge adaptive image steganography based on LSB matching revisited. *IEEE Trans. Inf. Forensics Secur.* **5**(2), 201–214 (2010).
7. Chakraborty, Soumendu, Anand Singh Jalal, and Charul Bhatnagar. "LSB based non blind predictive edge adaptive image steganography." *Multimedia Tools and Applications* 76 (2017): 7973–7987.
8. Bhatt, Santhoshi, et al. "Image steganography and visible watermarking using LSB extraction technique." *2015 IEEE 9th international conference on intelligent systems and control (ISCO)*. IEEE, 2015.
9. Dadgostar, H., and Fatemeh Afsari. "Image steganography based on interval-valued intuitionistic fuzzy edge detection and modified LSB." *Journal of information security and applications* 30 (2016): 94–104.
10. Lu, T.-C., Tseng, C.-Y. & Jhih-Huei, Wu. Dual imaging-based reversible hiding technique using LSB matching. *Signal Process.* **108**, 77–89 (2015).
11. Swain, Gandharba, and Saroj Kumar Lenka. "Pixel value differencing steganography using correlation of target pixel with neighboring pixels." *2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. IEEE, 2015.
12. Hussain, Mehdi, et al. "A data hiding scheme using parity-bit pixel value differencing and improved rightmost digit replacement." *Signal Processing: Image Communication* 50 (2017): 44–57.
13. Hosam, Osama, and Nadhir Ben Halima. "Adaptive block-based pixel value differencing steganography." *Security and Communication Networks* 9.18 (2016): 5036–5050.
14. Swain, G. A steganographic method combining LSB substitution and PVD in a block. *Procedia Computer Science* **85**, 39–44 (2016).
15. Kalita, Manashee, and Themrichon Tuithung. "A novel steganographic method using 8-neighboring PVD (8nPVD) and LSB substitution." *2016 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE, 2016.
16. Sahu, A. K. & Sahu, M. Digital image steganography and steganalysis: A journey of the past three decades. *Open Computer Science* **10**(1), 296–342. https://doi.org/10.1515/comp-2020-0136 (2020).
17. Savithri, G., et al. "Parallel implementation of RSA 2D-DCT steganography and chaotic 2D-DCT steganography." *Proceedings of International Conference on Computer Vision and Image Processing: CVIP 2016, Volume 1*. Springer Singapore, 2017.
18. Kumar, V. & Kumar, D. A modified DWT-based image steganography technique. *Multimedia Tools and Applications* **77**, 13279–13308 (2018).
19. Khashandarag, Asghar Shahrzad, et al. "An optimized color image steganography using LFSR and DFT techniques." *Advanced Research on Computer Education, Simulation and Modeling: International Conference, CESM 2011, Wuhan, China, June 18–19, 2011. Proceedings, Part II*. Springer Berlin Heidelberg, 2011.
20. Baluja, Shumeet. "Hiding images in plain sight: Deep steganography." *Advances in neural information processing systems* 30 (2017).
21. Duan, Xintao, et al. "Reversible image steganography scheme based on a U-Net structure." *IEEE Access* 7 (2019): 9314–9323.
22. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*. Springer international publishing, 2015.
23. C. Yu, Attention based data hiding with generative adversarial networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 1120–1128.
24. S.-P. Lu, R. Wang, T. Zhong, P.L. Rosin, Large-capacity image steganography based on invertible neural networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 10816–10825.
25. J. Jing, X. Deng, M. Xu, J. Wang, Z. Guan, HiNet: Deep image hiding by invertible network, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4733–4742.
26. Tan, J., Liao, X., Liu, J., Cao, Y. & Jiang, H. Channel attention image steganography with generative adversarial networks. *IEEE Trans. Netw. Sci. Eng.* **9**(2), 888–903 (2022).
27. F. Li, Y. Sheng, X. Zhang, C. Qin, iSCMIS: Spatial-channel attention based deep invertible network for multi-image steganography, IEEE Trans. Multimed. (2023) 1–15.
28. Available online: https://www.kaggle.com/datasets/imsparsh/flowers-dataset(accessed on 11 December 2024).
29. Available online: https://www.kaggle.com/datasets/muhammedtausif/rose-flowers(accessed on 11 December 2024).
30. Available online: https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection (accessed on 11 December 2024).
31. Available online: https://www.kaggle.com/datasets/sshikamaru/glaucoma-detection (accessed on 11 December 2024).
32. Zhu, Jiren, et al. "Hidden: Hiding data with deep networks." *Proceedings of the European conference on computer vision (ECCV)*. 2018.
33. Zhang, Kevin Alex, et al. "SteganoGAN: High-capacity image steganography with GANs." *arXiv preprint* arXiv:1901.03892 (2019).
34. Li, Yafeng, et al. "HCISNet: Higher-capacity invisible image steganographic network." *IET Image Processing* 15.13 (2021): 3332–3346.
35. Ambika, V. & Uplaonkar, D. S. Deep learning-based coverless image steganography on medical images shared via cloud. *Engineering Proceedings* **59**(1), 176 (2024).

Acknowledgements

The authors would like to acknowledge the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R757), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. Special appreciation is also extended to FIST-DST for providing the essential infrastructure and support that

greatly facilitated this research.The authors would like to thank Prince Sultan University for their valuable support. Sincere gratitude is also expressed to AICTE IDEA Lab, Amal Jyothi College of Engineering, Kanjirappally and APJ Abdul Kalam Technological University for providing their facilities and support towards the successful completion of this research.

Author contributions

Sherin Zafar, Kashish Ara Shakil, and Mudasir Ahmad Wani: Led the conceptualization and design of the steganography framework, developing the theoretical foundation of the model, formulating the research hypothesis, and drafting the initial manuscript. Bini M Issac and S N Kumar: Played significant role in the data collection, pre-processing, implementing and optimising architecture, evaluation of the system's performance, refining the manuscript and contributing to the critical review process.

Declarations

Competing interests

There is no competing interest.

Ethical approval

Not Applicable.

Additional information

Correspondence and requests for materials should be addressed to S.N.K., S.Z. or K.A.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.