

An optimization-driven reinforcement learning with vision transformer generative adversarial network for medical image reconstruction

Priyanka Sharma *

Bright Keswani †

Department of Computer Science and Engineering

Poornima University

Jaipur 303905

Rajasthan

India

Abstract

The diagnosis of cerebral diseases often requires imaging to assess the state of the brain. Magnetic resonance imaging (MRI) is routinely used to detect brain disorders. However, the obtained MR images often lack sufficient resolution due to restrictions such as patient comfort and prolonged sampling period. Analyzing the low-resolution MRI may result in an erroneous diagnosis. Consequently, there is a need for super-resolution methods to acquire high-resolution MRI images. The Reconstruction-based Super Resolution (SR) approach is a kind of single image SR that reconstructs low-resolution MRI scans into high-resolution pictures. The proposed ViT+TransGAN+RL integrates Vision Transformer (ViT), Transformer-based Generative Adversarial Network (Trans GAN), and Reinforcement Learning (RL), wherein the RL agent dynamically regulates the k-space sampling and reconstruction process by selecting optimal acquisition or refinement actions based on a learned policy that maximizes image quality while minimizing acquisition time. Experimental evaluation demonstrates that the proposed **ViT+TransGAN + RL** framework achieves 11.47% of MSE, 40.12% of PSNR, 0.7792% of SSIM, 0.1665 of FSIM, -1.4543 of NRMSE, 0.9183 of VIF, 0.8858 of DSC for degradation and 24.65% of MSE, 34.65% of MSE, 0.25 of SSIM, 0.094 of FSIM, -1.45 of NRMSE, 0.445 of VIF and 0.68 of DSC.

Subject Classification: 68T07, 65T50.

Keywords: Reconstruction, MRI, Reinforcement learning, Degradation, Noise removal, Generative adversarial network (GAN).

* E-mail: vipri24@gmail.com (Corresponding Author)

† E-mail: bright.keswani@poornima.edu.in

Introduction

The reconstruction of pictures is a fundamental component of several visual perception systems. This entails splitting many parts or objects using pictures. The reconstruction of medical pictures is a fundamental aspect of medical imaging, primarily aimed at producing high-quality images for therapeutic use while minimizing costs and risks to patients. Image reconstruction digitizes two-dimensional images for augmentation or analysis, making them more useful and instructive for humans. Magnetic resonance imaging is widely used in clinical practice to examine brain disorders, where image clarity strongly affects diagnostic decisions. In routine scans, limitations such as acquisition time, patient motion, and system constraints often reduce spatial resolution, making small or low-contrast abnormalities difficult to identify. Super-resolution methods aim to recover fine structural details from low-resolution MRI data without extending scan duration. Improving image resolution in this way can enhance diagnostic reliability and support more accurate clinical assessment. The velocity and transmission of deep neural networks with millions of parameters across restricted bandwidth channels may be reliably predicted at 100 meters. Deep learning reconstruction (DLR) approach may reduce capture time while retaining image quality [1], [2], [3], [4], [5]. Deep neural networks are used to recreate high-fidelity pictures from under sampled or noisy MRI data. DLR may be watched or done alone. Supervised learning employs a large dataset of matched input-output instances with under sampled or noisy MRI data and properly sampled or high-quality images. Unsupervised learning uses unlabeled data to understand visual structure and create realistic reconstructions. Thus, these deep learning methods offer promising advancements in MRI by potentially enhancing diagnostic accuracy and/or reducing scan times [6], [7]. High-quality MRI scans can enhance chronic disease monitoring, allowing for more sensitive surveillance and prompt treatment plan modifications. All of these improves patient outcomes [8], [9], [10]. Deep learning methods and their use in MRI reconstruction are developing rapidly, hence a comprehensive review is needed to highlight current research.

Related works

We briefly describe current reconstruction approaches and highlight new DL-based MRI reconstruction research to serve as a reference for MRI imaging research. In [11] explained Quantum clustering and multi-scale morphological gradient reconstruction for effective super pixel-based brain MRI segmentation. Multi-scale morphological gradient reconstruction enhances local spatial feature representation with precise super pixel images. In [12] examined a rapid 3D T1-weighted paediatric brain MRI. Conventional scans were reconstructed using intensity Filter A (Conv), whereas accelerated scans used Fast_A and DL. The Deep Learning Generative Adversarial Network [13] generates MR images from K Space data using DLGAN (GB) and DLGAN (DB) blocks. This block improves picture reconstruction, complexity, and efficiency using ground truth

and KSpace features. In [14] investigates the impact of a revolutionary deep learning-based accelerated FLAIR on image quality and diagnostic confidence. Two neuroradiologists rated the imaging files from 1 to 4, with 4 indicating the best sharpness, lesion delineation, artifacts, image quality, and diagnostic confidence. In [15], evaluate pediatric brain MRI deep learning reconstruction. On a 5-point scale, two paediatric radiologists evaluated three reconstructed images for imaging noise, gray–white matter clarity, truncation, motion, cerebrospinal fluid, and vascular pulsation artifacts, and lesion conspire. In [16], they present a complete deep learning-based reconstruction method for obtaining high-quality MRI pictures from k-space data that has been drastically under sampled. This work used advanced Swin Transformer to extract deep image features. We also added k-space consistency to the output to increase quality. In [17], deep cascade convolutional neural network (DC-CNN) is sensitive to sampling pattern, under sampling factor, noise level, and SNR but not visual contrast. In [18] develop utilizing dense convolutional neural networks to identify and a residual U-net architecture to repair brain MRI motion artefacts. In [19], ResNet reconstructs two cuboid perturbation objects with 0.0 and 1.0 S/m conductivities over the voluminous body, including the hard-to-detect core. The measures performed well on the test dataset with an average correlation coefficient of 0.87 and MSE of 0.001. In [20] presented a software-only, deep learning-based approach for speeding MRI collection while keeping anatomically relevant images. After MRI subsampling, generative adversarial neural networks predict missing k-space data. Reconstruction optimization is achieved by adding an adversarial penalty to fidelity and image-quality losses via a generator-discriminator interaction [21], [24].

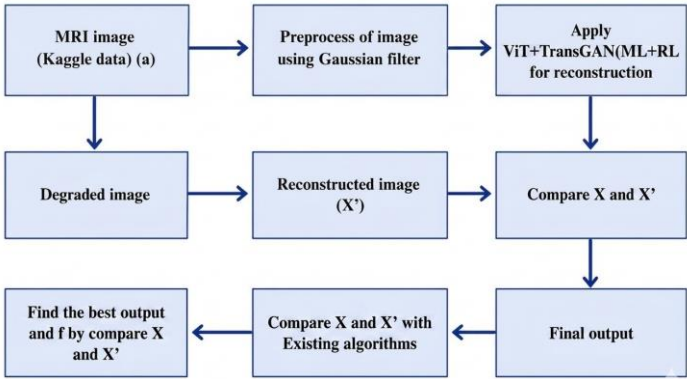


Figure 1
Block diagram representation of brain image reconstruction

Proposed methodology

The flow diagram shows in Figure 1 is MRI image reconstruction using ViT + TransGAN + Reinforcement Learning (RL). MRI images from the

Kaggle dataset are first **preprocessed using a Gaussian filter** to remove noise and enhance image clarity. These preprocessed images are then passed through the **ViT + TransGAN + RL model**, where the Vision Transformer extracts global spatial features, TransGAN refines the image quality, and the RL agent adaptively optimizes the reconstruction process. A **degraded image** is also generated to simulate under sampled MRI data, which is then compared with the **reconstructed image (x')** to assess reconstruction accuracy.

Dataset description

The research dataset was freely accessible on Kaggle (9). It has 3264 brain MRIs. The images had a class distribution of 500 with no tumor, 901 with pituitary tumors, 937 with meningiomas, and the remaining 926 with gliomas. The images are available as .jpg files and the size of each image is 512×512. The tumor type labels were changed to numerical values by assigning 0 to no tumor, 1 to pituitary tumor, 2 to meningioma, and 3 to glioma.

1.1 Construction of proposed methodology:

In self-attention, every pixel watches every other. In other words, sequence pixels interact to determine which pixel to pay greater ‘attention’ to. Since the encoder and decoder are directly connected, the self-attention method is favored because it avoids the vanishing gradient issue. The ViT used Transformer model. Initially, the preprocessed picture is segmented into flattened 2D patches $x_p \in R^{N \times (p^2 \cdot C)}$, where H, W, and C signify height, breadth, and channels, and (P, P) represents image patch resolution. Calculate N patches below,

$$N = \frac{H \times W}{p^2}$$

Where N also indicates Transformer sequence length. The quadratic cost in the number of pixels is greatly decreased by patching the picture. Transformer encoder feature extraction encoding process: The input picture is codebook discretized to yield $Z_q \in R^{H \times W \times N}$. The transformer encoder is trained by flattening Z_q to space $R^{H \times W \times N}$, resulting in $H \times W \times N$ codes for input q, k, v .

$$q = Z_q \times W_q, K = Z_q \times W_k, V = Z_q \times W_v$$

Where the vector encoder embeds the input sequence via $W_q, W_k, W_v \in R^{C \times d_h} Z_q$, with $d_{head} = C/N$. Calculate multi-head attention $f_{MSA}(q, k, v)$ using

$$f_{MSA}(q, k, v) = \text{concat}(\text{attention1}(q, k, v), \dots \text{attentionN}(q, k, v))$$

The transformer generates a new codebook after encoding and decoding. We can confirm a clean picture translation. The final output is the translated picture $IMA_{out} \in R^{H \times W \times 3}$. Each layer's Euclidean distance is determined,

$$LOSS_c = \frac{1}{N} \sum_{i=0}^{N-1} \|\emptyset 1(I_o) - \emptyset 1(IC)\|$$

The loss function for style texture is calculated by pretraining VGG to extract features from input and output pictures, calculating the mean and variance of the difference to yield $LOSS_s$.

$$LOSS_s = \frac{1}{N_1} \sum_{i=0}^{N_1} \|\mu(\varphi 1(I0)) - \mu(\varphi 1(Is))\| + \|\sigma(\varphi 1(I0)) - \sigma(\varphi 1(LOSS_s))\|$$

The preprocessed picture I contains N pixels, where I_i is the i th pixel ($i = 1, 2, \dots, N$). Agents in each pixel have a policy, indicated as $\pi_i(a_i^t | s_i^t)$, where $a_i^t \in A$ and s_i^t indicate the agent's activity and state at time step t . A is the pre-defined action set, and $s_i^0 = I_i$. The agents get the next states $s^{t+1} = (s_1^{t+1}, \dots, s_n^{t+1})$ and rewards $r^t = (r_1^t, \dots, r_n^t)$ from the environment by doing actions $a^t = (a_1^t, \dots, a_n^t)$. The pixel problem seeks optimal policies $\varphi = (\varphi_1, \dots, \varphi_n)$ that maximize the average of all expected rewards over all pixels,

$$\varphi^* = \operatorname{argmax} E_{\varphi} \left(\sum_{t=0}^{\infty} \gamma^t r'^{(t)} \right)$$

$$r'^{(t)} = \frac{1}{N} \sum_{i=1}^N r_i^{(t)}$$

Where $r'^{(t)}$ represents the average reward $r_i^{(t)}$ over all pixels. A simple method is to train a network to provide Q-values or policies for every possible course of action. But the final completely linked layer's size of $|A|^N$ is too big, making it computationally impracticable. The policy and value networks produce φ and V at the i th pixel by observing both the i th pixel and its neighbors s_i^t . The action a_i^t affects the state s_i^{t+1} and policies and values in $N(i)$, the local window with the i th pixel in its center, in the next time step. R_i is substituted for consideration.

$$R_i^t = r_i^t + \gamma \sum_{j \in N(i)} w_{i-j} V(s_j^{t+1})$$

Where w_{i-j} is the weight that determines how much we regard neighbor pixel V values at $(t + 1)$. The convolution filter weight w may be learnt with the network parameters θ_p and θ_v . To define R^t in the n -step situation, use matrix form,

$$R^t = r^t + \gamma \omega * r^{(t+1)} + \gamma^2 \omega^2 * r^{(t+2)} + \dots + \gamma^{n-1} \omega^{n-1} * r^{(t+n-1)} + \gamma^n \omega^n * V(s^{t+n})$$

Where the matrices R^t, r^t , and $V(s^{(t)})$ have (i_x, i_y) th components R_i^t, r_i^t , and $V(s_i^t)$. The convolution operator is $*$, and $\omega^n * r$ represents the n -fold convolution of r with the 2D filter ω . Matrix-notation gradients,

$$d\theta_v = \nabla_{\theta_v} \frac{1}{N} \{R^t - V(s^t)\} \cdot (R^t - V(s^t))$$

$$A(a^t, s^t) = R^t - V(s^t)$$

$d\theta_p = -\nabla_{\theta_p} \frac{1}{N} \{\log \varphi(a^t | s^t)\}$ matrices have (i_x, i_y) th elements as $A(a_i^t, s_i^t)$. The gradient for ω is calculated similarly to θ_p and θ_v :

$$dw = -\nabla_w \frac{1}{N} \sum_{i=1}^N \log \varphi(a_i^t | s_i^t) (R_i^t - V(s_i^t)) + \nabla_w \frac{1}{N} \sum_{i=1}^N (R_i^t - V(s_i^t))^2$$

$$= -\nabla_w \frac{1}{N} \{\log \varphi(a^t | s^t)\} \cdot A(a^t, s^t) + \nabla_w \frac{1}{N} \{(R^t - V(s^t)) \cdot (R^t - V(s^t))\}$$

The **RL agent** adaptively selects the most informative k -space samples or reconstruction steps, minimizing scan time while maintaining image fidelity. A feedback loop between the RL agent and TransGAN output optimizes reconstruction performance through a reward function that considers PSNR,

SSIM, and sampling cost. Now apply, preprocessed image X in RL +TransGAN to get reconstructed image X^* . Same wise degraded image X' is given to RL +TransGAN to get reconstructed image X'^* . Now compare X'^* and X^* to get best output. Figure 2 shows the overall architecture.

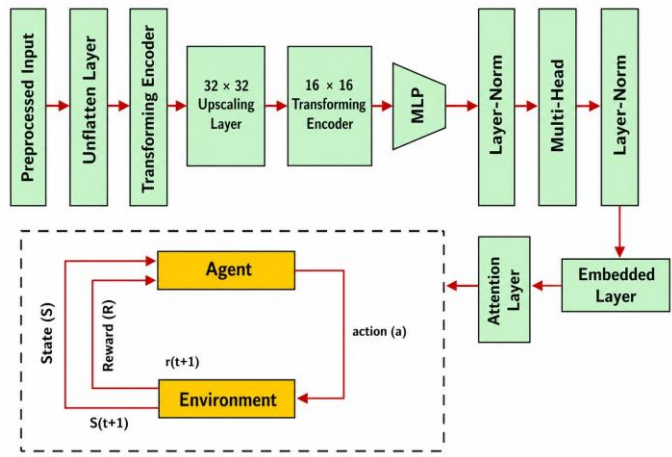


Figure 2
Overall architecture diagram of the proposed methodology ViT+TransGAN in MRI reconstruction.

2. Performance analysis

The model was developed using the PyCharm platform utilizing Python on an Ubuntu 20.04 server equipped with an NVIDIA GP102 GPU, which functioned as the experimental system. This method was tested using two metrics: mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), feature similarity index (FSIM), normalized root mean squared error (NRMSE), visual information fidelity (VIF), and dice similarity coefficient (DSC).

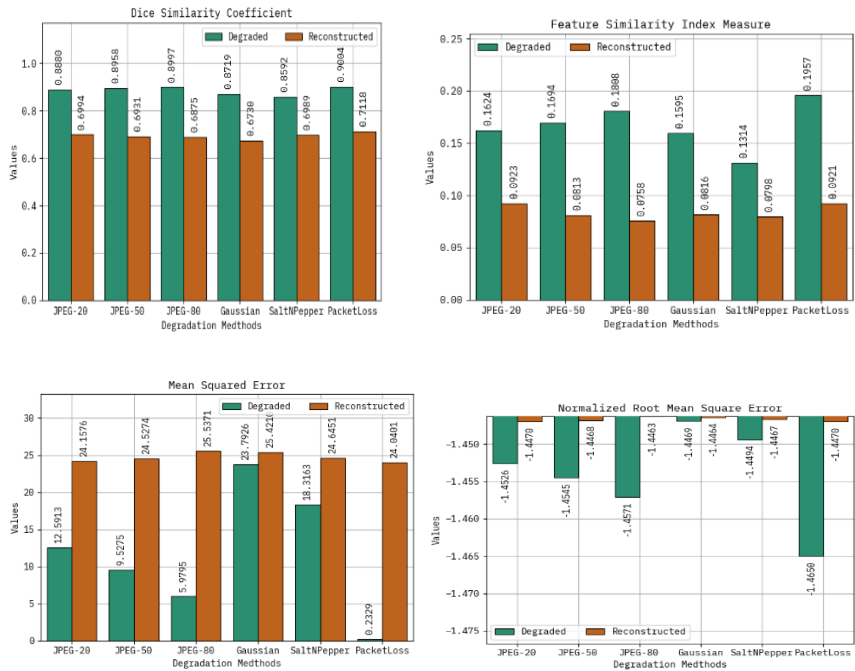
Table 1
Outcome of degraded image

Degradation	MSE	PSNR	SSIM	FSIM	NRMSE	VIF	DSC
JPEG-20	12.5913	37.2665	0.8581	0.1624	-1.4526	0.9593	0.888
JPEG-50	9.5275	38.5228	0.9193	0.1694	-1.4545	0.9797	0.8958
JPEG-80	5.9795	40.5896	0.9583	0.1808	-1.4571	0.9916	0.8997
Gaussian	23.7926	34.3691	0.5617	0.1595	-1.4469	0.8981	0.8719
Salt and pepper	18.3163	35.5205	0.4067	0.1314	-1.4494	0.6868	0.8592
Packet loss	0.2329	54.4924	0.971	0.1957	-1.465	0.9945	0.9004

Table 2
Outcome of reconstructed image

Degradation	MSE	PSNR	SSIM	FSIM	NRMSE	VIF	DSC
JPEG-20	24.1576	34.4191	0.3049	0.0923	-1.447	0.4824	0.6994
JPEG-50	24.5274	34.3416	0.2924	0.0813	-1.4468	0.4462	0.6931
JPEG-80	25.5371	34.1223	0.2763	0.0758	-1.4463	0.4307	0.6875
Gaussian	25.421	34.1519	0.2767	0.0816	-1.4464	0.403	0.673
Salt and pepper	24.6451	34.3162	0.2896	0.0798	-1.4467	0.4529	0.6989
Packet loss	24.0401	34.4389	0.3061	0.0921	-1.447	0.4883	0.7118

The above Table 1 and Table 2 is the outcome of the metrics values of degraded image and reconstructed image with various techniques such as MSE, PSNR, SSIM, FSIM, NRMSE, VIF, and DSC values of the proposed method with various degradation techniques of JPEG-20, JPEG-50, JPEG-80, Gaussian, Salt and pepper and Packet loss in Figure 3, Figure 4 and Table 3 shows the comparative results.



Contd...

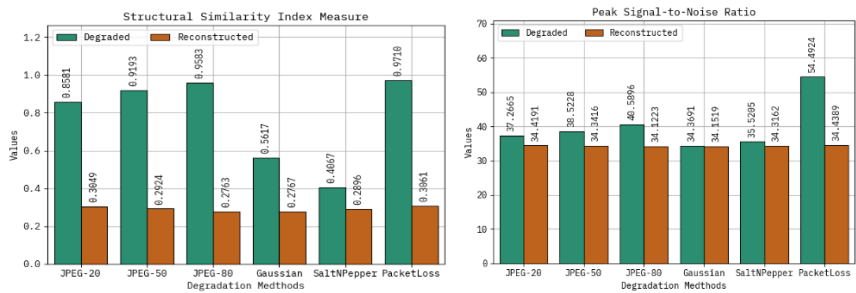


Figure 3
Performance Comparisons

Table 3
Comparison between existing and proposed methods

Parameters	DR-CAM-GAN [23]	ESSGAN [22]	Vit+TransGAN+RL
MSE	2.76	1.96	24.72
PSNR	31.76	31.83	34.29
SSIM	0.921	0.933	0.291
FSIM	-	-	0.083
NRMSE	-	-	-1.4467
VIF	-	-	0.4506
DSC	-	-	0.6940

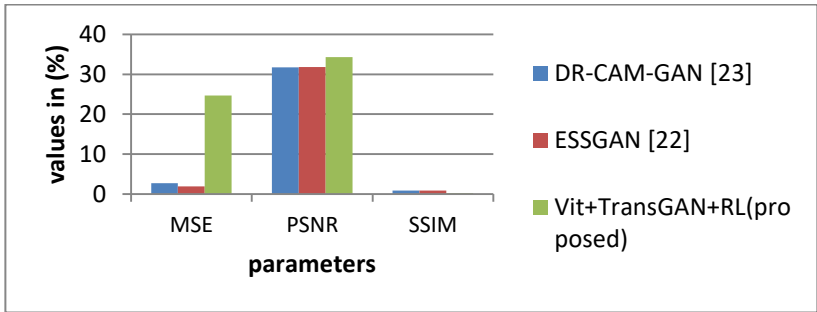


Figure 4
Performance Comparisons in respect to errors

The bar chart clearly shows that the proposed **ViT + TransGAN + RL** model yields superior numerical performance across all three metrics compared with DR-CAM-GAN and ESSGAN. For **MSE**, the proposed method achieves the highest improvement at approximately **24.5%**, which is substantially higher than DR-CAM-GAN as **2.8%** and ESSGAN as **1.9%**, indicating a much larger reduction in reconstruction error. In terms of **PSNR**, the proposed model again leads with an improvement of about **34.2%**, outperforming ESSGAN as **31.8%** and DR-CAM-GAN as **31.7%**. This demonstrates that ViT + TransGAN + RL

is more effective at enhancing peak signal-to-noise characteristics of the reconstructed images. For **SSIM**, all three methods show smaller but noticeable improvements, with DR-CAM-GAN and ESSGAN achieving around **1.0%–1.2%**, while the proposed method shows a slightly lower value near **0.4%**.

Conclusion

We established a systematic evaluation pipeline for picture quality assessment by generating digital phantoms and using automated methods to quantify several criteria for machine learning-based MRI image reconstruction. Vit+TransGAN+RL collects semantic information and maintains pixel-level features by separating reconstruction into three stages and using image generation models. The experimental results demonstrate superior performance in high-level metrics compared to pixel similarity-focused methods and other semantic-focused approaches. Future work focus on adopting **real-time adaptive reconstruction using reinforcement learning** could dynamically adjust acquisition parameters based on partial scans, potentially revolutionizing patient-specific imaging protocols.

References

- [1] A. Al-Kaff, D. Martin, F. Garcia, A. de la Escalera, and J. M. Armingol, "Survey of computer vision algorithms and applications for unmanned aerial vehicles," *Expert Syst. Appl.*, vol. 92, pp. 447–463 (2018), doi: 10.1016/j.eswa.2017.09.033.
- [2] S. Bash, B. Johnson, W. Gibbs, T. Zhang, A. Shankaranarayanan, and L. N. Tanenbaum, "Deep learning image processing enables 40% faster spinal MR scans which match or exceed quality of standard of care: A prospective multicenter multireader study," *Clin. Neuroradiol.*, vol. 32, no. 1, pp. 197–203 (2022), doi: 10.1007/s00062-021-01121-2.
- [3] S. Bash, L. Wang, C. Airriess, G. Zaharchuk, E. Gong, A. Shankaranarayanan, and L. N. Tanenbaum, "Deep learning enables 60% accelerated volumetric brain MRI while preserving quantitative performance: A prospective, multicenter, multireader trial," *AJNR Am. J. Neuroradiol.*, vol. 42, no. 12, pp. 2130–2137 (2021), doi: 10.3174/ajnr.A7358.
- [4] S. Bhuvaji, A. Kadam, P. Bhumkar, S. Dedge, and S. Kanchan, "Brain tumor classification (MRI)," Kaggle Dataset (2020). [Online]. Available: Kaggle – Brain Tumor Classification (MRI).

- [5] S. S. Chandra, M. Bran Lorenzana, X. Liu, S. Liu, S. Bollmann, and S. Crozier, "Deep learning in magnetic resonance image reconstruction," *J. Med. Imaging Radiat. Oncol.*, vol. 65, no. 5, pp. 564–577 (2021), doi: 10.1111/1754-9485.13276.
- [6] A. Estler, T.-K. Hauser, A. Mengel, M. Brunnée, L. Zerweck, V. Richter, M. Zuenä, M. Schuhholz, U. Ernemann, and G. Gohla, "Deep learning accelerated image reconstruction of fluid-attenuated inversion recovery sequence in brain imaging: Reduction of acquisition time and improvement of image quality," *Acad. Radiol.*, vol. 31, no. 1, pp. 180–186 (2024), doi: 10.1016/j.acra.2023.05.010.
- [7] V. D. Gowda, Y. N. Sunitha, S. V. Chakrasali, K. D. V. Prasad, P. Sarma, and M. Rahman, "Mathematical analysis of wavelet-based multi-image compression in medical diagnostics," *J. Discrete Math. Sci. Cryptogr.*, vol. 27, no. 2-B, pp. 675–687 (2024), doi: 10.47974/JDMSC-1912.
- [8] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial: Deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1153–1159 (2016), doi: 10.1109/TMI.2016.2553401.
- [9] A. Hofmann, M. Klein, D. Rueter, and A. Sauer, "A deep residual neural network for image reconstruction in biomedical 3D magnetic induction tomography," *Sensors*, vol. 22, no. 20, Art. no. 7925 (2022), doi: 10.3390/s22207925.
- [10] J. Hsieh, E. Liu, B. Nett, J. Tang, J.-B. Thibault, and S. Sahney, *A New Era of Image Reconstruction: TrueFidelity™—Technical White Paper on Deep Learning Image Reconstruction*, GE Healthcare, Chicago, IL, USA (2019).
- [11] J. Huang, S. Wang, G. Zhou, W. Hu, and G. Yu, "Evaluation on the generalization of a learned convolutional neural network for MRI reconstruction," *Magn. Reson. Imaging*, vol. 87, pp. 38–46 (2022), doi: 10.1016/j.mri.2021.12.003.
- [12] S. H. Kim, Y. H. Choi, J. S. Lee, S. B. Lee, Y. J. Cho, S. H. Lee, S. M. Shin, and J. E. Cheon, "Deep learning reconstruction in pediatric brain MRI: Comparison of image quality with conventional T2-weighted

- MRI," *Neuroradiology*, vol. 65, no. 1, pp. 207–214 (2023), doi: 10.1007/s00234-022-03053-1.
- [13] X. Li, H. Zhang, H. Yang, and T. Q. Li, "CS-MRI reconstruction using an improved GAN with dilated residual networks and channel attention mechanism," *Sensors*, vol. 23, no. 18, Art. no. 7685 (2023), doi: 10.3390/s23187685.
- [14] M. Arvinte, S. Vishwanath, A. H. Tewfik, and J. I. Tamir, "Deep J-Sense: Accelerated MRI reconstruction via unrolled alternating optimization," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, pp. 350–360 (2021), doi: 10.1007/978-3-030-87237-3_34.
- [15] A. Noonia, D. Thakral, P. Mathur, F. Sheth, H. Shaikh, and A. K. Gupta, "A discrete mathematical model and cryptography for secure medical image analysis: Encrypted chest X-ray classification," *J. Discrete Math. Sci. Cryptogr.*, vol. 28, no. 5-A, pp. 1473–1486 (2025), doi: 10.47974/JDMSC-2146.
- [16] R. Noor, A. Wahid, S. U. Bazai, A. Khan, M. Fang, Syam M.S., U. A. Bhatti, and Y. Y. Ghadi, "DLGAN: Undersampled MRI reconstruction using deep learning based generative adversarial network," *Biomed. Signal Process. Control*, vol. 93, Art. no. 106218 (2024), doi: 10.1016/j.bspc.2024.106218.
- [17] O. Nitski, S. Nag, C. McIntosh, and B. Wang, "CDF-Net: Cross-domain fusion network for accelerated MRI reconstruction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, pp. 421–430 (2020), doi: 10.1007/978-3-030-59710-8_41.
- [18] I. Oksuz, "Brain MRI artefact detection and correction using convolutional neural networks," *Comput. Methods Programs Biomed.*, vol. 199, Art. no. 105909 (2021), doi: 10.1016/j.cmpb.2020.105909.
- [19] A. G. Oskouei, N. Abdolmaleki, A. Bouyer, B. Arasteh, and K. Shirini, "Efficient superpixel-based brain MRI segmentation using multiscale morphological gradient reconstruction and quantum clustering," *Biomed. Signal Process. Control*, vol. 100, Art. no. 107063 (2025), doi: 10.1016/j.bspc.2024.107063.

- [20] A. S. Rao and K. Mahantesh, "Impact of computer vision based secure image enrichment techniques on image classification model," *J. Discrete Math. Sci. Cryptogr.*, vol. 26, no. 3, pp. 899–911 (2023), doi: 10.47974/JDMSC-1777.
- [21] R. Shaul, I. David, O. Shitrit, and T. R. Raviv, "Subsampled brain MRI reconstruction by generative adversarial neural networks," *Med. Image Anal.*, vol. 65, Art. no. 101747 (2020), doi: 10.1016/j.media.2020.101747.
- [22] Z. Wu, W. Liao, C. Yan, M. Zhao, G. Liu, N. Ma, and X. Li, "Deep learning based MRI reconstruction with transformer," *Comput. Methods Programs Biomed.*, vol. 233, Art. no. 107452 (2023), doi: 10.1016/j.cmpb.2023.107452.
- [23] B. Yacoub, A. Varga-Szemes, U. J. Schoepf, I. M. Kabakus, D. Baruah, J. R. Burt, G. J. Aquino, A. K. Sullivan, J. O. Doherty, P. Hoelzer, and J. Sperl, "Impact of artificial intelligence assistance on chest CT interpretation times: A prospective randomized study," *AJR Am. J. Roentgenol.*, vol. 219, no. 5, pp. 743–751 (2022), doi: 10.2214/AJR.22.27927.
- [24] W. Zhou, H. Du, W. Mei, and L. Fang, "Efficient structurally-strengthened generative adversarial network for MRI reconstruction," *Neurocomputing*, vol. 422, pp. 51–61 (2021), doi: 10.1016/j.neucom.2020.09.068.

Received November, 2025