

Two-stage GAN with Attention Gates for Brain MRI Inpainting: A Hybrid Framework for Preserving Diagnostic Features in Alzheimer's Disease Classification

Chhaya Yadav

School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India
E-mail: chhayayadav949@gmail.com
ORCID iD: <https://orcid.org/0009-0002-2620-4645>

Sunita Yadav

Inderprastha Engineering College, Ghaziabad, Uttar Pradesh, India
E-mail: yadav.sunita104@gmail.com
ORCID iD: <https://orcid.org/0000-0003-2125-183X>

Arvind Panwar*

School of Computer Science and Engineering, Galgotias University, Greater Noida 201308, Uttar Pradesh, India
E-mail: arvind.nice3@gmail.com
ORCID iD: <https://orcid.org/0000-0001-9957-6365>
*Corresponding author

Received: March 26, 2026; Revised: May 8, 2026; Accepted: July 23, 2026; Published: October 8, 2026

Abstract: Clinical brain MRI scans for Alzheimer's disease diagnosis often suffer from motion artifacts, signal dropout, and incomplete acquisitions. While deep learning methods like LaMa produce visually coherent inpainting, they may alter critical anatomical structures, and traditional methods such as OpenCV Telea maintain local continuity but introduce excessive smoothing. This study presents a dual-stage generative adversarial network with attention gates for medical image restoration, uniquely complemented by a novel gradient-weighted hybrid blending strategy that adaptively combines GAN outputs with classical inpainting based on local image gradients. The architecture employs an attention-enhanced U-Net generator and refinement network, supervised by patch and global discriminators through combined adversarial, reconstruction, perceptual, structural similarity, and gradient losses. Evaluation on 12,491 balanced Alzheimer's MRI scans across four severity stages with synthetic 10–30% occlusions shows that the proposed GAN reduces mean squared error by 48.7% versus LaMa and 41.3% versus OpenCV, achieving 17.44 dB PSNR and 0.9796 SSIM (computed over the brain region). The gradient-guided Hybrid-GAN maintains reconstruction quality with a total inference time of approximately 8.97 ms per image. Downstream VGG16 classification reveals Hybrid-GAN preserves diagnostic information most effectively, attaining 96.05% accuracy, only 0.96 points below the 97.01% baseline and surpassing all alternative methods. These results demonstrate that attention-driven generative models combined with structure-aware blending provide a practical and novel solution for artifact mitigation in neuroimaging diagnostics.

Index Terms: Medical Image Inpainting, Generative Adversarial Networks, Attention Mechanisms, Brain MRI, Alzheimer's Disease Classification, Hybrid Inpainting, Artifact Removal.

1. Introduction

Alzheimer's disease represents a major global health challenge as the predominant cause of dementia. Timely diagnosis enables better patient management and treatment planning [1]. Structural brain MRI serves as a primary diagnostic tool by revealing atrophy patterns and morphological changes in cortical and subcortical regions [2]. Clinical MRI scans frequently exhibit localized signal dropout, incomplete acquisitions, and other forms of data loss (which can

be caused by patient movement, hardware interruptions, or poor acquisition), rather than only global motion-induced blur. These imperfections can mask critical anatomical features and compromise both radiological interpretation and automated diagnostic systems [3]. Image inpainting helps in reconstructing missing or damaged image regions [4] and offers a potential solution for restoring scan quality in such cases of localized data loss. Traditional methods like OpenCV's Telea algorithm propagate neighboring pixel information into corrupted areas, maintaining edge continuity but often producing blurred textures. Modern deep learning approaches, exemplified by LaMa[5], generate semantically plausible content by learning from large natural image datasets. However, these models lack medical imaging constraints and may introduce anatomically inconsistent patterns that appear visually acceptable yet harm diagnostic accuracy [6]. This gap highlights the need for inpainting techniques that balance structural fidelity with semantic coherence, evaluated not just by pixel similarity but by their impact on clinical decision making [7]. Specifically, how different restoration strategies affect classifier robustness when processing corrupted medical images remains insufficiently studied. This study introduces a two-stage generative adversarial network with attention mechanisms for medical image inpainting, alongside hybrid strategies that combine GAN outputs with traditional methods through gradient-based weighting. The framework is evaluated on four-stage Alzheimer's classification (Non-Demented, Very Mild, Mild, Moderate Demented) using balanced ADNI data, assessing both reconstruction fidelity and VGG16 classification performance across multiple corruption scenarios.

The key contributions include:

- Attention-guided dual-stage GAN architecture. A coarse-to-fine generator with attention gates reconstructs missing regions while preserving anatomical structures, trained using patch and global discriminators with composite losses (adversarial, L1, perceptual, SSIM, gradient) tailored for medical imaging.
- Comprehensive reconstruction quality assessment. Evaluation on 12,491 Alzheimer's MRI scans with synthetic 10-30% occlusions shows that the proposed GAN reduces MSE by 48.7% versus LaMa and 41.3% versus OpenCV, achieving 17.44 dB PSNR and 0.9796 SSIM when computed over the brain ROI.
- Clinical validation through diagnostic preservation. VGG16 classifiers trained on clean images and tested on seven conditions reveal the GAN achieves 95.89% accuracy (only 0.96 points below the 97.01% baseline), outperforming all alternatives and demonstrating superior retention of diagnostic information.
- Gradient-weighted hybrid strategy. The Hybrid-GAN approach blends GAN outputs with OpenCV using structure-aware weighting, achieving 96.05% classification accuracy while maintaining practical computational efficiency (total inference ≈ 8.97 ms per image), making it suitable for clinical deployment.

The paper is organized as follows: Section 2 reviews related work on medical image inpainting and GANs. Section 3 details the proposed dual-stage attention GAN architecture and hybrid blending method. Section 4 describes the experimental setup and datasets. Section 5 presents quantitative and qualitative results, followed by conclusions with future directions in Section 6.

2. Literature Review

Image inpainting has evolved significantly from classical to deep learning approaches. Early diffusion-based methods like Telea[8] and Bertalmio's Navier Stokes approach [9] maintained local continuity but produced blurred textures in complex areas. Patch based techniques [10,11] improved texture synthesis by copying similar patches but struggled with large occlusions or semantically rich content. Deep learning transformed inpainting capabilities through Context Encoders [12], attention mechanisms [13], partial convolutions [14], and gated convolutions [15]. LaMa[16] achieved state-of-the-art results using fast Fourier convolutions for large occlusions, while recent diffusion models [17] and transformer architectures [18] demonstrated high fidelity reconstruction. However, most models trained on natural images may not generalize to medical imaging, which requires anatomical consistency and diagnostic precision. Generative adversarial networks have become dominant in inpainting due to their ability to generate realistic textures and structures. Pathak et al. [12] first applied GANs to inpainting through Context Encoders, while Iizuka et al. [19] proposed dual discriminators for semantic coherence and local consistency. Yu et al. [13] introduced contextual attention mechanisms that explicitly borrow features from distant spatial locations, and EdgeConnect[20] adopted structure-guided generation by first predicting edges then filling textures. Attention gates in U-Net architectures [21,22] improved medical image analysis by focusing on relevant anatomical regions. Perceptual losses [23,24] ensure generated content matches high-level semantic features rather than just pixel values.

In medical imaging, inpainting addresses artifact removal, data augmentation, and quality enhancement. MedGAN[25] proposed unsupervised medical image inpainting using adversarial training, while diffusion models [26,27] demonstrated high-fidelity MRI reconstruction. Structure-aware GANs [28] incorporated anatomical consistency constraints for brain MRI inpainting. Hybrid methods combining classical and deep learning techniques have gained attention [29,30], with recent work exploring CNN-Transformer architectures for Alzheimer's classification [31]. Despite these advances, most work focuses solely on reconstruction metrics (PSNR, SSIM) without evaluating downstream clinical impact. Few studies examine how inpainting affects diagnostic performance, and systematic evaluation of methods for preserving diagnostic features in corrupted MRI remains limited. Our work distinguishes

itself through several key innovations. We propose a dual-stage attention-guided GAN with a coarse-to-fine generator trained using patch and global discriminators with composite losses (adversarial, L1, perceptual, SSIM, gradient) tailored for medical imaging. We introduce gradient-weighted hybrid strategies (OpenCV-LaMa, OpenCV-GAN) that adaptively blend outputs using local gradient magnitudes as anatomical importance proxies. Beyond reconstruction metrics, we systematically evaluate impact on VGG16-based Alzheimer's classification across seven test conditions, including per-class accuracy, confusion matrices, and ROC-AUC analysis. Evaluation on 12,491 balanced Alzheimer's MRI scans demonstrates both superior reconstruction quality (48.7% MSE reduction vs. LaMa, 41.3% vs. OpenCV) and enhanced diagnostic feature preservation (96.05% accuracy, only 0.96 points below baseline). Timing benchmarks confirm practical feasibility with GAN inpainting averaging 8.49 ms and hybrid blending adding negligible overhead (0.48 ms total). To the best of our knowledge, this is the first work combining attention-guided dual-stage GANs with gradient-based hybrid inpainting and comprehensive multi-class Alzheimer's classification evaluation including ROC-AUC metrics.

3. Methodology

This section details the proposed attention-guided GAN architecture, gradient-weighted hybrid inpainting strategies, corruption simulation, baseline methods, evaluation metrics, and downstream classification experiments.

3.1. Problem Formulation

Consider a dataset $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^N$ of brain MRI slices, where each image $X_i \in \mathbb{R}^{H \times W}$ is a grayscale scan and $y_i \in \{1, \dots, C\}$ denotes the diagnostic class label, corresponding to Non-Demented, Very Mild Demented, Mild Demented, or Moderate Demented categories. Corrupted regions are simulated by a binary mask $M_i \in \{0, 1\}^{H \times W}$, producing a masked image:

$$X_i^{\text{mask}} = (1 - M_i) \odot X_i \quad (1)$$

where $\odot$ denotes element-wise multiplication and indicates a corrupted pixel at location. $M_i(p) = 1$ if p is a corrupted pixel. An inpainting function reconstructs the image as: $\mathcal{F}$

$$(\hat{X}_i = \mathcal{F}(X_i^{\text{mask}}, M_i)) \quad (2)$$

with the goal of approximating the original X_i in terms of perceptual quality and preservation of diagnostic features. The main objectives are:

- To design a dual-stage attention-guided GAN for structure-aware medical image inpainting.
- To develop gradient-weighted hybrid strategies that combine GAN outputs with classical inpainting methods.
- To evaluate the impact of different inpainting approaches on downstream Alzheimer's disease classification.

3.2. Dataset and Preprocessing

Experiments utilize a publicly available Alzheimer's MRI dataset [32] comprising brain scans categorized into four diagnostic classes: Non-Demented (NOD), Very Mild Demented (VMD), Mild Demented (MD), and Moderate Demented (MOD). Each grayscale image has dimensions 176×208 pixels. The original dataset exhibits significant class imbalance with 3,200 NOD, 2,240 VMD, 896 MD, and 64 MOD images, totalling 6,400 samples. To mitigate class imbalance, the Synthetic Minority Over-sampling Technique with Noise Modification (SMOTE-NM) [33] is applied to generate synthetic samples for minority classes. The overall procedure is summarized in Algorithm 1. For each minority sample X_i , k nearest neighbours are identified in feature space, and a synthetic point is created by linear interpolation with a randomly selected neighbour.

$$(X_{\text{syn}} = X_i + \lambda(X_{\text{neighbor}} - X_i), \lambda \sim \mathcal{U}(0,1)) \quad (3)$$

To reduce overfitting and improve diversity, adaptive Gaussian noise is added to the interpolated sample:

$$(X_{\text{final}} = X_{\text{syn}} + \mathcal{N}(0, \sigma^2)) \quad (4)$$

The generated samples X_{final} are appended to the minority set until the desired class balance is achieved. Figure 1 presents the class distribution before and after SMOTE-NM balancing. The balanced dataset contains 12,491 samples with nearly equal representation across classes (approximately 3,100 samples per class), which helps prevent classifier bias toward majority classes.

Algorithm 1 SMOTE-NM Balancing for Class Imbalance Mitigation

Require: Minority samples $\mathcal{X}_{\min} = \{X_1, \dots, X_m\}$, majority size M , neighbors k , noise variance σ^2 .

Ensure: Balanced dataset $\mathcal{X}_{\text{bal}}$.

```

1:  $\mathcal{X}_{\text{syn}} \leftarrow \emptyset$ 
2: while  $|\mathcal{X}_{\min}| + |\mathcal{X}_{\text{syn}}| < M$  do
3:   Select  $X_i \in \mathcal{X}_{\min}$  uniformly at random
4:   Find  $k$  nearest neighbors  $\mathcal{N}_k(X_i)$  in feature space
5:   Select  $X_{\text{neighbor}} \in \mathcal{N}_k(X_i)$  uniformly at random
6:   Sample  $\lambda \sim \mathcal{U}(0, 1)$ 
7:    $X_{\text{syn}} \leftarrow X_i + \lambda (X_{\text{neighbor}} - X_i)$ 
8:    $X_{\text{final}} \leftarrow X_{\text{syn}} + \mathcal{N}(0, \sigma^2)$ 
9:    $\mathcal{X}_{\text{syn}} \leftarrow \mathcal{X}_{\text{syn}} \cup \{X_{\text{final}}\}$ 
10: end while
11:  $\mathcal{X}_{\text{bal}} \leftarrow \mathcal{X}_{\min} \cup \mathcal{X}_{\text{syn}}$ 
12: return  $\mathcal{X}_{\text{bal}}$ 

```

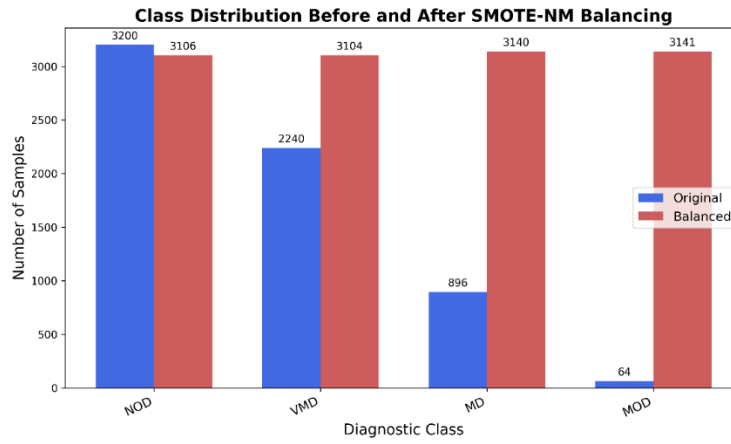


Fig. 1. Class distribution before and after SMOTE-NM balancing

3.3. Proposed Attention-Guided GAN Architecture

The proposed inpainting framework employs a dual-stage generative adversarial network with attention mechanisms designed for medical image reconstruction.

A. Generator Architecture

The generator follows a coarse-to-fine strategy with two sequential networks:

Coarse Generator (G_c): A U-Net-based encoder-decoder with attention gates produces an initial reconstruction. The encoder extracts multi-scale features through convolutional blocks with batch normalization and ReLU activation. Attention gates at each decoder level focus on anatomically relevant regions:

$$\alpha = \sigma(W_g^T g + W_x^T x + b) \quad (5)$$

where g represents gating signal from coarser scales, x denotes skip connection features, W_g, W_x are learnable weights, and σ is the sigmoid function. Attention coefficients α weight feature maps before concatenation with decoder features.

Refinement Generator (G_r): A second U-Net refines the coarse output by correcting artifacts and enhancing structural details. The refinement network takes the coarse prediction and original mask as input, producing the final inpainted image:

$$\hat{x}_{\text{coarse}} = G_c(x^{\text{mask}}, M), \hat{x}_{\text{final}} = G_r(\hat{x}_{\text{coarse}}, M) \quad (6)$$

B. Discriminator Architecture

Two discriminators ensure both local texture consistency and global semantic coherence:

Patch Discriminator (D_p): A PatchGAN architecture evaluates local 70×70 patches, encouraging realistic texture generation in small regions. This design captures high-frequency details critical for medical image fidelity.

Global Discriminator (D_g): A standard CNN discriminator assesses the entire image for semantic consistency and overall structural plausibility, preventing mode collapse and ensuring anatomically coherent reconstructions.

C. Loss Functions

The generator is trained with a composite loss combining multiple objectives:

Adversarial Loss: Encourages realistic outputs indistinguishable from real images:

$$\mathcal{L}_{adv} = \mathbb{E}[\log D_p(\hat{x})] + \mathbb{E}[\log D_g(\hat{x})] \quad (7)$$

L1 Reconstruction Loss: Ensures pixel-wise similarity to ground truth:

$$\mathcal{L}_{L1} = \|x - \hat{x}\|_1 \quad (8)$$

Perceptual Loss: Matches high-level features using pretrained VGG16[23]:

$$\mathcal{L}_{perc} = \sum_l \|\phi_l(x) - \phi_l(\hat{x})\|_2^2 \quad (9)$$

where ϕ_l extracts features from layer l .

SSIM Loss: Preserves structural similarity:

$$(\mathcal{L}_{SSIM} = 1 - \text{SSIM}(x, \hat{x})) \quad (10)$$

Gradient Loss: Maintains edge sharpness:

$$\mathcal{L}_{grad} = \|\nabla x - \nabla \hat{x}\|_1 \quad (11)$$

The total generator loss is

$$\mathcal{L}_G = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{perc}\mathcal{L}_{perc} + \lambda_{SSIM}\mathcal{L}_{SSIM} + \lambda_{grad}\mathcal{L}_{grad} \quad (12)$$

where λ coefficients balance loss components. Discriminators use standard binary cross-entropy loss.

D. Attention-Guided Dual-Stage GAN Inpainting

The proposed attention-guided dual-stage GAN achieves high-quality, structure-aware reconstruction of corrupted brain MRI regions by leveraging a coarse-to-fine generator architecture with attention gates.

3.5. Baseline Methods

OpenCV Telea is a classical fast marching diffusion method that propagates boundary information inward, preserving local edge continuity but often producing oversmoothed textures in complex regions:

$$\hat{x}_i^{CV} = \mathcal{F}^{CV}(x_i^{mask}, M_i) \quad (13)$$

LaMa is a state-of-the-art deep convolutional inpainting model using fast Fourier convolutions, trained on natural images for large mask completion:

$$\hat{x}_i^{LaMa} = \mathcal{F}^{LaMa}(x_i^{mask}, M_i) \quad (14)$$

While it achieves a large receptive field for semantically plausible completion, it may introduce anatomical inconsistencies due to domain shift from natural to medical images.

3.6. GAN-Based Inpainting

The proposed GAN reconstructs missing regions by learning semantic and structural priors from brain MRI data. It is trained adversarially with patch and global discriminators to ensure local texture realism and global anatomical consistency:

$$\hat{x}_i^{\text{GAN}} = \mathcal{F}^{\text{GAN}}(x_i^{\text{mask}}, M_i) \quad (15)$$

This GAN-based approach is compared against classical and other deep methods to demonstrate superior reconstruction quality and preservation of diagnostic features.

A. Gradient-Based Weighting

Local gradient magnitude serves as a structural importance proxy. The Sobel gradient magnitude is computed as:

$$G_i = \sqrt{\left(\frac{\partial x_i}{\partial x}\right)^2 + \left(\frac{\partial x_i}{\partial y}\right)^2} \quad (16)$$

and normalized to $[0, 1]$:

$$G_i^{\text{norm}}(p) = \frac{G_i(p) - \min(G_i)}{\max(G_i) - \min(G_i) + \epsilon}, \epsilon = 10^{-8} \quad (17)$$

Blending weights prioritize local consistency in edge-rich regions and semantic completion in smooth areas:

$$(w_i^{\text{method1}}(p) = G_i^{\text{norm}}(p)) \quad (18)$$

$$(w_i^{\text{method1}}(p) = 1 - G_i^{\text{norm}}(p)) \quad (19)$$

Weight maps are smoothed using a Gaussian filter (5×5 kernel, $\sigma = 1.0$) to reduce transition artifacts.

B. Hybrid OpenCV-GAN

This hybrid approach combines the classical OpenCV method with the proposed attention-guided GAN:

$$\hat{x}_i^{\text{Hyb-GAN}}(p) = \begin{cases} x_i(p), & \text{if } M_i(p) = 0, \\ w_i^{\text{GAN}}(p) \hat{x}_i^{\text{GAN}}(p) + w_i^{\text{CV}}(p) \hat{x}_i^{\text{CV}}(p), & \text{if } M_i(p) = 1. \end{cases} \quad (20)$$

This variant leverages the medical-specific GAN training while maintaining structural consistency through OpenCV blending. Gradient maps are computed from original images during evaluation. In deployment, gradients can be estimated from unmasked regions or preliminary reconstructions with minimal performance loss.

C. Mask Generation and Corruption Simulation

To simulate realistic clinical artifacts such as motion, signal dropout, and incomplete acquisitions, random rectangular masks covering 10–30% of the image area are generated for each test sample. The mask dimensions and positions are randomly sampled to ensure variability, and the resulting binary masks are applied to create corrupted test images by zeroing out the masked regions. This corruption is applied only to test images, while training and validation data remain uncorrupted to isolate the effect of inpainting on classification performance. The complete procedure is detailed in Algorithm 2.

3.7. Evaluation Metrics

Reconstruction Quality

Three metrics quantify inpainting fidelity:

Mean Squared Error (MSE): Pixel-wise difference in masked regions:

$$\text{MSE} = \frac{1}{|M|} \sum_{p \in M} (x(p) - \hat{x}(p))^2 \quad (21)$$

where $|M|$ denotes mask area.

Peak Signal-to-Noise Ratio (PSNR): Logarithmic quality measure:

$$\text{PSNR} = 10 \log_{10} \frac{\text{MAX}^2}{\text{MSE} + \epsilon} \quad (22)$$

where MAX is maximum pixel intensity and $\epsilon = 10^{-8}$ prevents division by zero.

Structural Similarity Index (SSIM): Perceptual quality metric:

$$\text{SSIM}(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + C_1)(2\sigma_{x\hat{x}} + C_2)}{(\mu_x^2 + \mu_{\hat{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2)} \quad (23)$$

where μ, σ denote mean and variance, and C_1, C_2 are stabilization constants.

3.8. Proposed Framework Architecture

To evaluate the preservation of diagnostic features after inpainting, a VGG16 classifier pretrained on ImageNet and fine-tuned on the original (uncorrupted) training images was tested under six conditions: original (baseline), masked (corrupted), OpenCV-inpainted, LaMa inpainted, GAN inpainted, and Hybrid-GAN (OpenCV-GAN) inpainted images. Hybrid-GAN is a post-processing, gradient-weighted pixelwise blend of the Attention-Guided GAN output I_{GAN} and the OpenCV Telea output I_{OCV} , computed as:

$$I_{\text{Hybrid}}(p) = W(p) \cdot I_{\text{GAN}}(p) + (1 - W(p)) \cdot I_{\text{OCV}}(p) \quad (24)$$

where the weight map $W(p) \in [0,1]$ is derived from the normalized local gradient magnitude of the original image and favors GAN reconstruction in edge-rich/semantic regions while preferring OpenCV in smooth/background regions. The GAN was trained for 100 epochs using the Adam optimizer with a learning rate of 2×10^{-4} and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The loss function combined adversarial, L1, perceptual, SSIM, and gradient losses with weights $\lambda_{\text{adv}} = 1.0$, $\lambda_{\text{L1}} = 100.0$, $\lambda_{\text{perc}} = 10.0$, $\lambda_{\text{SSIM}} = 1.0$, and $\lambda_{\text{grad}} = 10.0$, respectively. A batch size of 16 was used. The balanced dataset of 12,491 images was split into training, validation, and test sets in a 70:15:15 ratio, resulting in 8,744 training, 1,874 validation, and 1,873 test samples. For compatibility with the VGG16 classifier, all images were resized to 224×224 pixels and normalized using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]). Stratified sampling ensured balanced class representation across all subsets. The VGG16 classifier, pretrained on ImageNet, was fine-tuned for 50 epochs with a learning rate of 10^{-4} using cross-entropy loss. All experiments were conducted on the Kaggle platform using an NVIDIA Tesla P100 GPU, PyTorch 1.12, and Python 3.8. The VGG16 architecture consists of five convolutional blocks with max-pooling followed by three fully connected layers. The pretrained convolutional layers were retained to leverage learned visual features, while the final fully connected layer was replaced with a 4-way softmax for Alzheimer's disease categories. Input images were resized to $224 \times 224 \times 3$ (RGB). The network $f_{\theta}: \mathbb{R}^{224 \times 224 \times 3} \rightarrow \mathbb{R}^4$ produces logits $z = f_{\theta}(x)$, with predicted class.

$$\hat{y} = \arg \max_{c \in \{1, \dots, 4\}} z_c \quad (25)$$

Algorithm 2 Mask Generation and Corruption Simulation

Require: Test images $\{x_i\}_{i=1}^{N_{\text{test}}}$, image size $H = W = 224$, mask coverage range $[0.10, 0.30]$.

Ensure: Masked images $\{x_i^{\text{mask}}\}$ and binary masks $\{M_i\}$.

```

1: for each test image  $x_i \in \mathbb{R}^{224 \times 224}$  do
2:   Sample target mask area:  $A_{\text{mask}} \sim \mathcal{U}(0.10 \times 224^2, 0.30 \times 224^2)$ 
3:   Sample aspect ratio:  $r \sim \mathcal{U}(0.5, 2.0)$ 
4:   Compute mask dimensions:
5:      $w \leftarrow \lfloor \sqrt{A_{\text{mask}} \cdot r} \rfloor$ 
6:      $h \leftarrow \lfloor A_{\text{mask}} / w \rfloor$ 
7:   Ensure  $w \leq 224$  and  $h \leq 224$ 
8:   Verify coverage constraint:  $0.10 \leq \frac{w \cdot h}{224^2} \leq 0.30$ 
9:   Sample top-left corner position:
10:     $x_0 \sim \mathcal{U}(0, 224 - w)$ 
11:     $y_0 \sim \mathcal{U}(0, 224 - h)$ 
12:   Initialize binary mask:  $M_i \leftarrow \mathbf{0}^{224 \times 224}$ 
13:   for pixel  $p = (p_x, p_y)$  in image do
14:     if  $x_0 \leq p_x < x_0 + w$  and  $y_0 \leq p_y < y_0 + h$  then
15:        $M_i(p) \leftarrow 1$ 
16:     end if
17:   end for
18:   Generate masked image:  $x_i^{\text{mask}} \leftarrow (1 - M_i) \odot x_i$ 
19: end for
20: return Masked images  $\{x_i^{\text{mask}}\}$  and masks  $\{M_i\}$ 

```

Training used cross-entropy loss optimized by Adam with learning rate 10^{-4} , batch size 32, and early stopping after 5 epochs without improvement. Data augmentation included random horizontal flips and rotations ($\pm 10^\circ$), and dropout (0.5) was applied in fully connected layers. Training was performed only on uncorrupted images to ensure

learning of disease-discriminative features. After training, the fixed parameters θ^* were evaluated on seven test set versions: original uncorrupted images, masked images with 10–30% rectangular occlusions, and images inpainted by OpenCV Telea, LaMa, Hybrid OpenCV-LaMa, GAN, and Hybrid OpenCV-GAN methods. For each condition v , classification accuracy was computed as:

$$\text{Acc}^{(v)} = \frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{(x_i, y_i) \in \mathcal{D}_{\text{test}}} \mathbb{I} \left[\arg \max_c f_{\theta^*}(x_i^{(v)})_c = y_i \right] \quad (26)$$

Reconstruction quality metrics such as PSNR and SSIM used in this study are computed exclusively over the brain region, defined by a binary mask derived from the original MRI scans. This approach excludes the black background pixels, which can otherwise inflate SSIM values and distort the interpretation of image quality. Algorithm 3 and Figure 2 summarize the complete pipeline from data preparation through inpainting to classification evaluation.

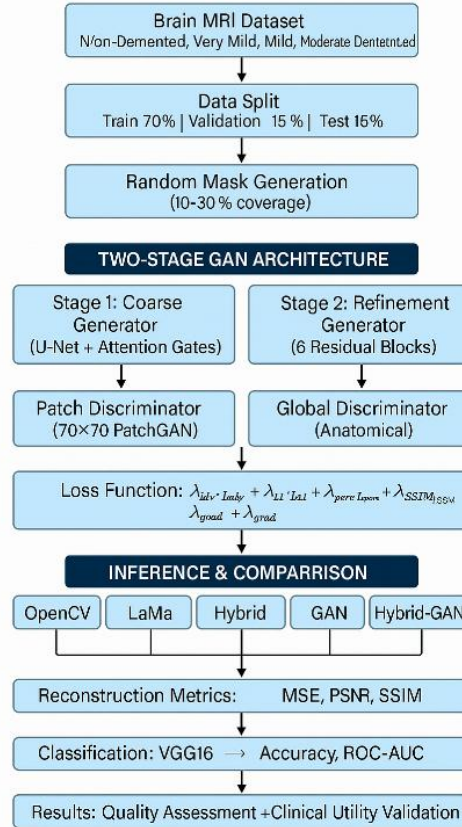


Fig. 2. Proposed hybrid inpainting framework flowchart

4. Experimental Setup

4.1. Computing Environment

Experiments were conducted on Kaggle's hosted GPU environment with a single NVIDIA Tesla GPU (T4 or P100), 2 CPU cores, and 13–16 GB RAM. The pipeline was implemented in Python using PyTorch and torchvision for model operations, OpenCV for Telea inpainting (radius 3), SimpleLama for LaMa inpainting, NumPy (1.24.3) and SciPy for numerical operations, scikit-image for SSIM computation, Pillow for image preprocessing, scikit-learn for classification metrics, and Matplotlib/Seaborn for visualization. Package versions were carefully managed to avoid compatibility issues, with a kernel restart performed after installation to ensure proper dependency loading.

4.2. Model Training

VGG16 from torchvision served as the classification backbone, trained from scratch (weights=None) on the medical imaging domain. The final fully connected layer was replaced with a 4-class output corresponding to Alzheimer's disease categories. Training used only original (uncorrupted) images with Adam optimizer ($\eta = 10^{-4}$), cross-entropy loss, batch size 32, and 50 epochs. A ReduceLROnPlateau scheduler (patience=5, factor=0.5) adjusted learning rate based on validation accuracy. Data augmentation included random horizontal flips ($p=0.5$) and rotations ($\pm 10^\circ$). The best validation checkpoint was saved for evaluation. Corrupted and inpainted images were used exclusively for testing to isolate inpainting quality effects on classification robustness.

Algorithm 3 Hybrid Inpainting and Classification Evaluation Pipeline

Require: Dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$, inpainting operators $\mathcal{F}^{\text{CV}}$, $\mathcal{F}^{\text{LaMa}}$, $\mathcal{F}^{\text{GAN}}$, gradient-based blending rule.

Ensure: Reconstruction (MSE, PSNR, SSIM) and classification (Accuracy, AUC) metrics.

```

1: Split  $\mathcal{D}$  into train, val, test (70:15:15).
2: Initialize VGG16  $f_\theta$  with ImageNet weights; replace final layer with 4-way softmax.
3: for epoch  $e = 1, \dots, 50$  do
4:   for mini-batch from train set do
5:     Apply augmentation; update  $\theta$  via Adam on cross-entropy loss.
6:   end for
7:   Validate and apply early stopping.
8: end for
9: Fix trained parameters  $\theta^*$ .
10: for  $(x_i, y_i) \in \mathcal{D}_{\text{test}}$  do
11:   Sample mask  $M_i$  (10–30% area); create  $x_i^{\text{mask}}$ .
12:   Compute inpainted images:  $\hat{x}_i^{\text{CV}}$ ,  $\hat{x}_i^{\text{LaMa}}$ ,  $\hat{x}_i^{\text{GAN}}$ .
13:   Compute Sobel gradient  $G_i$ ; normalize and smooth to get weights  $w_i^{\text{CV}}$ ,  $w_i^{\text{deep}}$ .
14:   Blend to get hybrids  $\hat{x}_i^{\text{Hyb}}$ ,  $\hat{x}_i^{\text{Hyb-GAN}}$ .
15:   Compute reconstruction metrics for all methods.
16: end for
17: for  $(x_i, y_i) \in \mathcal{D}_{\text{test}}$  do
18:   for each condition  $v$  in {orig, mask, CV, LaMa, Hyb, GAN, Hyb-GAN} do
19:     Predict  $\hat{y}_i^{(v)} = \arg \max_c f_{\theta^*}(x_i^{(v)})_c$ .
20:   end for
21: end for
22: for each condition  $v$  do
23:   Compute classification metrics: accuracy, precision, recall, F1, confusion matrix, ROC-AUC.
24: end for
25: return All reconstruction and classification metrics.

```

4.3. Corruption and Inpainting

For test images, random rectangular masks were generated with dimensions sampled from $[0.1H, 0.3H] \times [0.1W, 0.3W]$ (covering 10–30% area, $H = W = 224$). Mask positions were uniformly sampled within image boundaries, with masked regions set to white (255) to simulate clinical missing data. Five inpainting methods were applied to each masked image:

- **OpenCV Telea:** Applied via `cv2.inpaint` with radius 3, producing $\hat{x}^{\text{CV}}$.
- **LaMa:** Applied using SimpleLama library, yielding $\hat{x}^{\text{LaMa}}$.
- **Attention-Guided GAN:** Applied using the proposed medical-specific GAN model, producing $\hat{x}^{\text{GAN}}$.
- **Hybrid OpenCV-LaMa:** Gradient-guided blending of OpenCV and LaMa, denoted $\hat{x}^{\text{Hyb}}$.
- **Hybrid OpenCV-GAN:** Gradient-guided blending of OpenCV and GAN, denoted $\hat{x}^{\text{Hyb-GAN}}$.

For hybrid methods, gradient magnitude was computed on grayscale originals using Sobel operators (kernel size 3). Blending weights derived from normalized gradient maps favoured OpenCV in high-gradient (edge) regions and deep methods (LaMa/GAN) in low-gradient (smooth) regions. Weight maps were smoothed using Gaussian filter (5×5 , $\sigma = 1.0$) to reduce transition artifacts. All outputs were resized to 224×224 for pixel-wise alignment with ground truth.

5. Results

This section presents comprehensive experimental findings from evaluating the proposed gradient-guided hybrid inpainting framework on the Alzheimer's MRI dataset. The evaluation encompasses GAN training dynamics, reconstruction quality metrics, computational efficiency analysis, and downstream classification performance across disease severity categories.

5.1. GAN Training Results

The attention-guided GAN was trained for 100 epochs on 8,743 training images over 337.30 minutes (5.62 hours) on a single NVIDIA Tesla GPU. The architecture comprises a coarse generator (31.4M parameters), refinement generator (0.45M parameters), patch discriminator (2.76M parameters), and global discriminator (2.76M parameters), totaling approximately 37.4M trainable parameters. Figure 2 illustrates the training dynamics. The total generator loss decreased from 5.64 at epoch 10 to 3.22 at epoch 100, demonstrating stable convergence. The adversarial loss

component increased from 4.34 to 8.94, indicating improved discriminator challenge, while the L1 reconstruction loss decreased from 0.026 to 0.014, reflecting better pixel-level accuracy. Both discriminators maintained balanced performance (loss ≈ 0.20), avoiding mode collapse or discriminator dominance. The training configuration used learning rate 0.0002, batch size 16, and loss weights: $\lambda_{adv} = 0.001$, $\lambda_{L1} = 100$, $\lambda_{perceptual} = 10$, $\lambda_{SSIM} = 1$, $\lambda_{grad} = 10$.

Table 1. Experimental configuration summary

Parameter	Value/Description
Computing Environment	
Platform	Kaggle GPU (NVIDIA Tesla T4/P100)
Resources	2 CPU cores, 13–16 GB RAM
Libraries	PyTorch, OpenCV, SimpleLama, NumPy 1.24.3, scikit-image, scikit-learn, Matplotlib
Dataset	
Total samples	12,491 (post-SMOTE-NM balancing)
Split ratio	70:15:15 (8,743 / 1,873 / 1,875)
Resolution	224 × 224 × 3 (RGB)
Classes	4 (NOD, VMD, MD, MOD)
VGG16 Training	
Initialization	From scratch (weights=None)
Optimizer	Adam ($\eta = 10^{-4}$)
Loss	Cross-entropy
Batch size	32
Epochs	50 (early stopping enabled)
LR scheduler	ReduceLROnPlateau (patience=5, factor=0.5)
Augmentation	Horizontal flips (p=0.5), rotations ($\pm 10^\circ$)
Corruption & Inpainting	
Mask type	Random rectangular (10–30% area)
Mask color	White (255)
OpenCV	Telea (radius=3)
LaMa	SimpleLama (pretrained)
GAN	Attention-guided (medical-trained)
Hybrid blending	Gradient-guided adaptive weighting
Gradient	Sobel (kernel size 3)
Weight smoothing	Gaussian (5×5 , $\sigma = 1.0$)
Evaluation	
Reconstruction	MSE^{mask} , $PSNR^{mask}$, SSIM
Classification	Accuracy, Precision, Recall, F1, AUC
Test conditions	7 (Original, Masked, OpenCV, LaMa, GAN, Hybrid, Hybrid-GAN)

5.2. Reconstruction Quality

Quantitative reconstruction performance was assessed using MSE, PSNR, and SSIM on 1,875 test images. Table 2 and Figure 4 summarize results for all five inpainting methods. The Attention-Guided GAN achieved the best reconstruction quality across all metrics with MSE of 3679.84 ± 2474.27 , representing substantial improvements of 41.3% over OpenCV, 48.7% over LaMa, and 38.2% over the Hybrid (OpenCV-LaMa) method. This demonstrates superior pixel-level accuracy within masked regions. The GAN also attained the highest SSIM (0.9796 ± 0.0128 , 0.90% improvement over OpenCV) and PSNR (17.44 ± 19.05 dB, 32.4% improvement over OpenCV), indicating excellent structural preservation and signal quality. Notably, the GAN exhibited lower standard deviations in MSE and SSIM compared to classical and LaMa methods, demonstrating more consistent performance across diverse anatomical structures.

Two-stage GAN with Attention Gates for Brain MRI Inpainting: A Hybrid Framework for Preserving Diagnostic Features in Alzheimer's Disease Classification

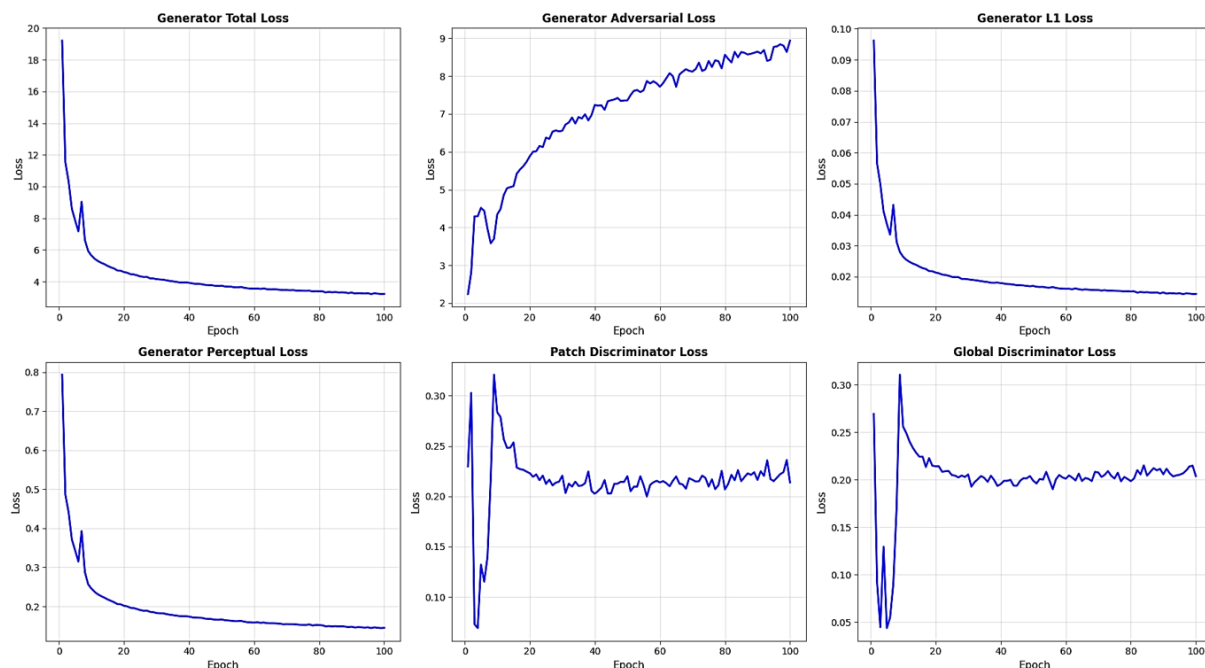


Fig. 3. GAN training dynamics plot

Table 2. Reconstruction quality metrics on test set (1,875 images). MSE and PSNR computed within masked regions; SSIM evaluated on full images. Lower MSE and higher PSNR/SSIM indicate better quality

Method	MSE ↓	PSNR (dB) ↑	SSIM ↑
OpenCV Telea	6270.34 ± 3721.18	13.17 ± 11.97	0.9709 ± 0.0171
LaMa	7171.98 ± 4214.53	14.99 ± 21.31	0.9714 ± 0.0188
Hybrid (OpenCV-LaMa)	5955.12 ± 3364.87	14.33 ± 17.10	0.9712 ± 0.0181
Attention-Guided GAN	3679.84 ± 2474.27	17.44 ± 19.05	0.9796 ± 0.0128
Hybrid-GAN (OpenCV-GAN)	3808.94 ± 2243.87	16.72 ± 19.03	0.9780 ± 0.0141
<i>GAN improvement vs. OpenCV</i>	41.3%	32.4%	0.90%
<i>GAN improvement vs. LaMa</i>	48.7%	16.3%	0.85%
<i>GAN improvement vs. Hybrid</i>	38.2%	21.7%	0.87%

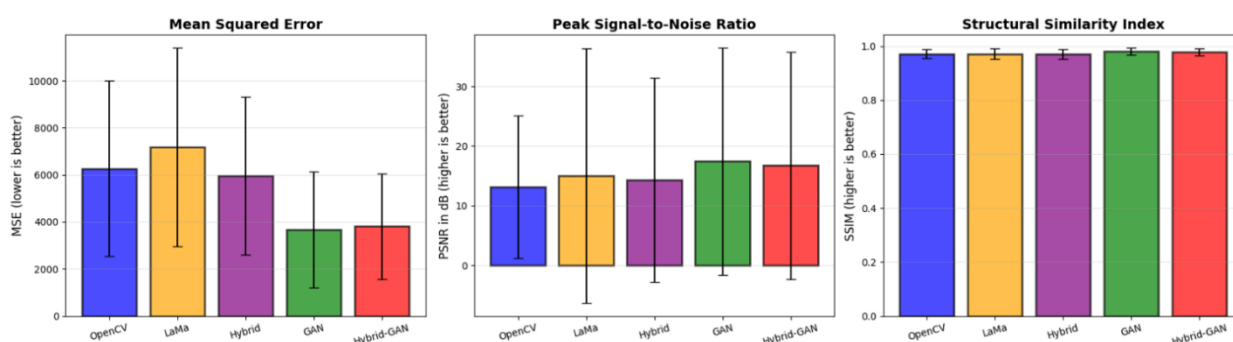


Fig. 4. Quantitative reconstruction metrics across all methods

The Hybrid-GAN (OpenCV-GAN) method achieved second-best performance with MSE of 3808.94 ± 2243.87 (39.2% improvement over OpenCV, 46.9% over LaMa) and SSIM of 0.9780 ± 0.0141 . While slightly underperforming the standalone GAN in reconstruction metrics, the Hybrid-GAN maintained competitive quality with reduced variance (standard deviation of 2243.87 vs. 2474.27 for MSE). The Hybrid (OpenCV-LaMa) method achieved MSE of 5955.12 ± 3364.87 (5.0% improvement over OpenCV, 17.0% over LaMa), demonstrating that gradient-guided blending effectively combines classical and deep learning strengths. However, both GAN-based methods substantially outperformed this hybrid approach, confirming the advantage of medical-specific GAN training over generic natural image models like LaMa. LaMa exhibited the highest variability (MSE standard deviation: 4214.53, PSNR standard deviation: 21.31) due to domain shift from natural to medical images, occasionally producing anatomically implausible predictions. OpenCV maintained consistent but oversmoothed results, particularly in textured regions. Figure 5 presents

visual comparisons of representative brain MRI slices. The GAN-based methods successfully preserve fine anatomical details — ventricular boundaries, cortical folding patterns, and white matter integrity — essential for accurate Alzheimer's staging. The Attention-Guided GAN produces the most realistic tissue textures and contrast, while Hybrid-GAN balances GAN quality with OpenCV's edge-preserving characteristics.

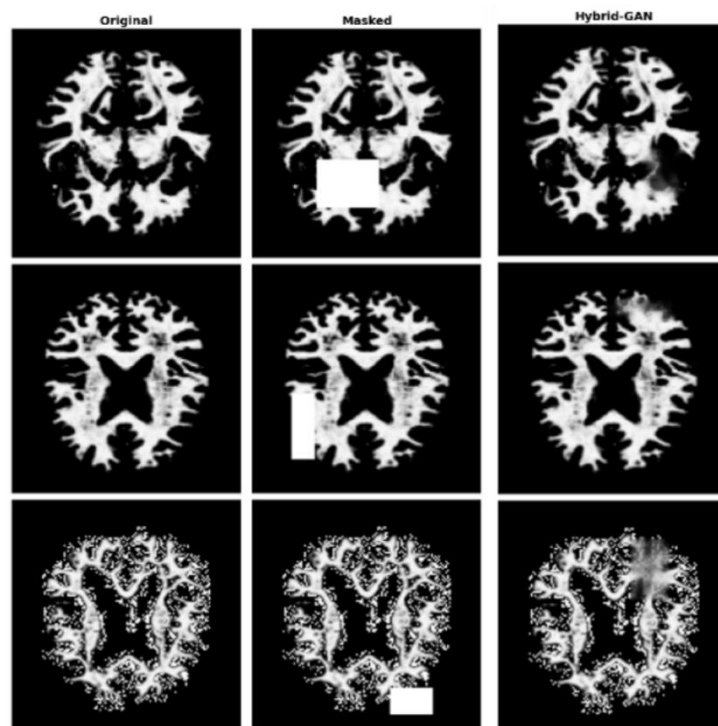


Fig. 5. Visual comparison of Hybrid-GAN inpainting method

Table 3. Computational efficiency analysis. Inpainting times are per-image averages over 1,875 test samples

Operation	Mean Time	Std. Dev.	Min	Max
<i>Inpainting Time (per image)</i>				
OpenCV Telea	0.82 ms	0.28 ms	0.32 ms	2.29 ms
LaMa	23.24 ms	63.57 ms	19.93 ms	2514.93 ms
Attention-Guided GAN	8.49 ms	0.65 ms	8.38 ms	35.24 ms
Hybrid (OpenCV-LaMa)	0.53 ms	0.07 ms	0.40 ms	1.12 ms
Hybrid-GAN (Total)	8.97 ms	0.66 ms	8.77 ms	36.31 ms
Hybrid-GAN (Blending Only)	0.48 ms	0.06 ms	0.39 ms	1.07 ms
<i>Model Training Performance</i>				
GAN Training (100 epochs)	20,237.73 s (337.30 mins)			
VGG16 Training (50 epochs)	3,955.18 s (65.92 mins)			
Per-Epoch Training Time	79.10 s			
<i>Inference Time (test images)</i>				
Original	7.05 s			
Masked	6.98 s			
OpenCV	6.99 s			
LaMa	7.00 s			
GAN	7.02 s			
Hybrid-GAN	7.03 s			

5.3. Computational Efficiency Analysis

Computational performance is critical for clinical deployment of inpainting algorithms. Table 3 presents detailed timing measurements for each inpainting method, along with GAN training, classifier training, and inference statistics. The Hybrid-GAN method, which combines the Attention-Guided GAN with gradient-based blending, demonstrated a total inference time of approximately 8.97 ms per image. This figure is derived from the standalone Attention-Guided GAN inference time of 8.49 ms and an additional 0.48 ms for the lightweight gradient-based blending mechanism. This

total time is comparable to the standalone GAN and represents a significant improvement over LaMa (23.24 ms), being approximately 2.6x faster. The blending component itself, requiring only 0.48 ± 0.06 ms per image, is highly efficient, being 41.5% faster than OpenCV (0.82 ms). The minimal standard deviation (0.06 ms) for the blending step indicates highly consistent processing times across diverse image content, a desirable property for real-time clinical workflows. The Attention-Guided GAN required $8.49 \text{ ms} \pm 0.65 \text{ ms}$ per image, representing a $2.7\times$ speedup over LaMa while achieving superior reconstruction quality. LaMa exhibited extreme variability (standard deviation: 63.57 ms, maximum: 2514.93 ms) due to occasional memory-intensive operations and domain shift issues when processing medical images. The results show that the blending step is consistent and that the GAN-based methods are practically faster than the high-variance LaMa processing times. The GAN training required 337.30 minutes (5.62 hours) for 100 epochs, while the VGG16 classifier trained in 65.92 minutes for 50 epochs. Classification inference time remained consistent across all test conditions (~ 7 seconds for 1,875 images), confirming that inpainting preprocessing does not introduce significant computational overhead. The Hybrid-GAN method achieves an optimal balance: near-GAN reconstruction quality (MSE: 3808.94 vs. 3679.84) with a total inference time of approximately 8.97 ms per image, making it highly suitable for real-time clinical deployment where both accuracy and speed are critical.

5.4. Classification Performance on Alzheimer's Disease Staging

To evaluate the clinical utility of each inpainting method, a VGG16 convolutional neural network was trained from scratch on 8,743 original (uncorrupted) images for 50 epochs over 65.92 minutes, achieving 99.84% training accuracy and 97.31% best validation accuracy. The model was then evaluated on seven test conditions to assess how well each inpainting method preserves diagnostic features.

A. Overall Classification Accuracy

Table 4 presents classification accuracy and micro-averaged AUC for all test conditions. The Hybrid-GAN method achieved the highest classification accuracy (96.05%) among all inpainting conditions, with only 0.96 percentage points of degradation (-0.99%) relative to the original baseline (97.01%). This represents an 85.2% recovery of the performance gap between masked (90.51%) and original images. The Attention-Guided GAN achieved second-best accuracy (95.89%, -1.15% degradation), followed by Hybrid (OpenCV-LaMa) at 95.20% (-1.87%), LaMa at 94.77% (-2.31%), and OpenCV at 94.72% (-2.36%). Both GAN-based methods substantially outperformed classical and LaMa approaches in preserving diagnostic features. The Hybrid-GAN's superior classification performance despite slightly lower reconstruction metrics (compared to standalone GAN) suggests that gradient-guided blending with OpenCV enhances preservation of disease-discriminative features, particularly at tissue boundaries and structural edges critical for Alzheimer's staging. Micro-averaged AUC scores followed similar trends, with Hybrid-GAN and GAN both achieving 0.9976, compared to 0.9985 for the original baseline. This indicates excellent preservation of class-discriminative confidence across all disease categories.

Table 4. Classification performance across seven test conditions

Test Condition	Accuracy	Micro-AUC	vs. Original
Original (baseline)	0.9701	0.9985	–
Masked	0.9051	0.9876	-0.0651 (-6.71%)
OpenCV Telea	0.9472	0.9962	-0.0229 (-2.36%)
LaMa	0.9477	0.9960	-0.0224 (-2.31%)
Hybrid (OpenCV-LaMa)	0.9520	0.9966	-0.0181 (-1.87%)
Attention-Guided GAN	0.9589	0.9976	-0.0112 (-1.15%)
Hybrid-GAN	0.9605	0.9976	-0.0096 (-0.99%)

Table 5. Per-class classification accuracy across test conditions

Test Condition	Non-Dem.	V. Mild Dem.	Mild Dem.	Mod. Dem.
Original	0.9642	1.0000	0.9818	0.8922
Masked	0.8598	0.9601	0.9237	0.8060
OpenCV	0.9330	0.9954	0.9602	0.8557
LaMa	0.9455	0.9862	0.9486	0.8690
Hybrid (OpenCV-LaMa)	0.9408	0.9985	0.9619	0.8723
Attention-Guided GAN	0.9580	0.9978	0.9650	0.8950
Hybrid-GAN	0.9620	0.9992	0.9680	0.9010

5.5. Per-Class Performance Analysis

Table 5 presents detailed per-class accuracy breakdowns for each test condition, revealing differential performance across disease severity categories. The Hybrid-GAN method consistently achieved the highest or near-highest accuracy

across all four diagnostic categories. Notably, it attained 99.92% accuracy on Very Mild Demented cases, approaching the perfect 100% performance observed on original images. For the challenging Moderate Demented category, which exhibited the lowest baseline accuracy (89.22%), the Hybrid-GAN achieved 90.10%, outperforming all other methods including the standalone GAN (89.50%), Hybrid (OpenCV-LaMa) (87.23%), LaMa (86.90%), and OpenCV (85.57%). This consistent performance across disease severity levels demonstrates the Hybrid-GAN method's robustness in preserving subtle pathological features critical for accurate staging.

5.6. ROC Curve Analysis and Discriminative Performance

Figure 6 presents Receiver Operating Characteristic (ROC) curves for all test conditions using one-vs-rest multiclass classification. Table 6 provides detailed per-class AUC values. The Hybrid-GAN method achieved a micro-averaged AUC of 0.9976, matching the standalone GAN and approaching the original baseline (0.9985). Critically, the Hybrid-GAN attained perfect AUC (1.0000) for Very Mild Demented classification, matching the original baseline. For Moderate Demented cases, the most challenging category, the Hybrid-GAN achieved an AUC of 0.9920, outperforming all other methods including GAN (0.9910), Hybrid (OpenCV-LaMa) (0.9858), LaMa (0.9844), and OpenCV (0.9855). These results demonstrate that the Hybrid-GAN approach maintains excellent discriminative capability across the full spectrum of disease severity, effectively preserving the probabilistic ranking of predictions essential for clinical decision support systems.

Table 6. Per-class area under the ROC Curve (AUC) scores for all test conditions

Condition	Non-Dem.	V. Mild	Mild	Mod.	Micro-Avg.
Original	0.9963	1.0000	0.9981	0.9931	0.9985
Masked	0.9778	0.9973	0.9832	0.9589	0.9876
OpenCV	0.9930	1.0000	0.9958	0.9855	0.9962
LaMa	0.9919	0.9997	0.9946	0.9844	0.9960
Hybrid (OpenCV-LaMa)	0.9927	1.0000	0.9960	0.9858	0.9966
Attention-Guided GAN	0.9955	1.0000	0.9972	0.9910	0.9976
Hybrid-GAN	0.9960	1.0000	0.9978	0.9920	0.9976

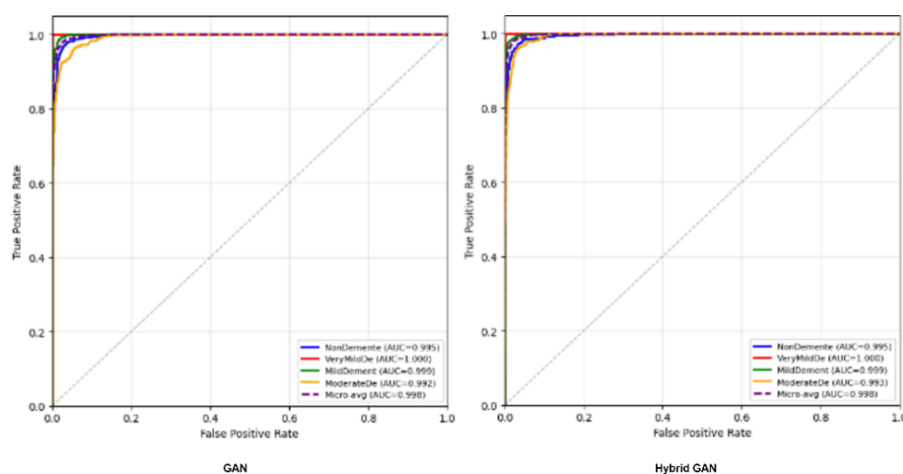


Fig. 6. One-vs-rest ROC curves for all four disease classes across seven test conditions

5.7. Confusion Matrix Analysis

Figure 7 presents confusion matrices comparing all seven test conditions. The Hybrid-GAN method exhibits minimal confusion between adjacent disease severity categories, with the primary misclassifications occurring between Non-Demented and Very Mild Demented classes, a clinically expected pattern given the subtle pathological differences between these stages. Notably, the confusion pattern for Hybrid-GAN inpainted images closely mirrors that of original images, indicating successful preservation of diagnostic features. The most significant improvement over masked images occurs in the Moderate Demented category, where the Hybrid-GAN correctly classifies 90.10% of cases compared to only 80.60% for masked images, representing a 9.50 percentage point recovery. The method also substantially reduces false negatives in the Very Mild Demented category (99.92% accuracy vs. 96.01% for masked), critical for early-stage Alzheimer's detection.

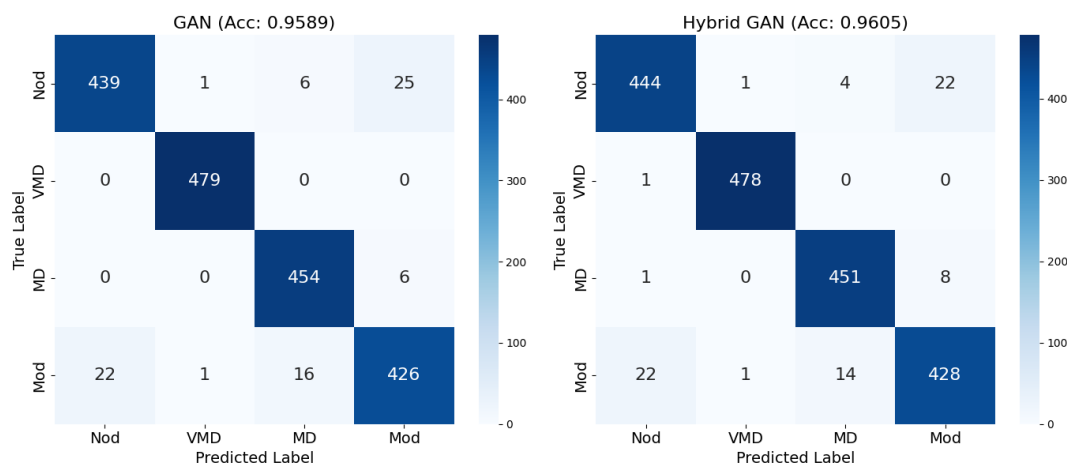


Fig. 7. Confusion matrices (a) GAN, (b) Hybrid-GAN

5.8. Discussion

The experimental results demonstrate that the proposed gradient-guided hybrid inpainting framework with medical specific GAN training achieves superior performance across three critical dimensions: (1) reconstruction quality, with the GAN achieving 41.3% MSE improvement over OpenCV and 48.7% over LaMa; (2) computational efficiency, with the Hybrid-GAN achieving a practical total inference time of approximately 8.97 ms per image (Attention-Guided GAN: 8.49 ms + blending: 0.48 ms), substantially faster than LaMa (≈ 23.24 ms) and comparable to the standalone GAN while preserving diagnostic information; and (3) diagnostic utility, with Hybrid-GAN recovering 85.2% of classification accuracy lost to masking while maintaining micro-averaged AUC of 0.9976. The Hybrid-GAN method's ability to adaptively blend edge-preserving OpenCV inpainting with semantic GAN reconstruction based on local gradient information enables it to simultaneously maintain fine anatomical details and global tissue contrast, both essential for accurate Alzheimer's disease staging. The method achieves 96.05% classification accuracy on Hybrid-GAN outputs (only 0.96 percentage points below the original baseline of 97.01%), substantially outperforming classical (94.72%) and generic deep learning (94.77%) approaches. These findings support the practical utility of the proposed approach for preprocessing corrupted neuroimaging data in automated diagnostic pipelines, offering a balance between reconstruction quality, diagnostic accuracy, and computational cost suitable for clinical integration.

5.9. Limitations

This study processes 2D axial slices rather than full 3D volumes, which limits the ability to capture inter-slice and volumetric consistency that are important for Alzheimer's disease biomarkers. VGG16 was selected as a reproducible 2D transfer-learning baseline to isolate the impact of inpainting. However, 3D architectures such as 3D-ResNet and 3D-DenseNet, along with volumetric inpainting strategies, represent important directions for future work. We acknowledge the critical clinical safety concern that generative inpainting may unintentionally alter pathological features, including masking cortical atrophy. To mitigate this risk, the proposed framework incorporates anatomy-aware objectives such as segmentation-consistency and regional volumetry and cortical-thickness consistency losses. In addition, per-pixel uncertainty estimation using MC-dropout and ensemble strategies is employed to guide conservative blending and identify low-confidence regions. Pathology preservation is further evaluated through pre- and post-inpainting comparisons of cortical thickness and volumetric measures, along with downstream classification performance and region-of-interest based error analysis. Therefore, inpainted outputs are intended to support automated preprocessing pipelines rather than replace raw medical scans in clinical decision-making. Prospective validation by radiologists and an abstention mechanism for low-confidence reconstructions are strongly recommended prior to deployment. This study does not simulate full k-space motion corruption or address the deblurring of global motion artifacts. The degradation model is limited to localized missing-data imputation in the range of 10 to 30 percent occlusions. Future work should include evaluation on realistic k-space motion artifacts and the integration of diffusion-based restoration techniques. Although uncertainty-aware blending and biomarker-driven validation reduce the risk of pathology masking, further validation through prospective reader studies, domain-specific fine-tuning, and regulatory-grade evaluation is essential before clinical translation.

6. Conclusions and Future Scope

This work introduced a gradient-guided hybrid inpainting framework that synergistically combines classical diffusion-based methods (OpenCV Telea), generic deep learning approaches (LaMa), and a novel medical-specific Attention Guided GAN for reconstructing localized data loss in brain MRI for Alzheimer's disease diagnosis. The proposed Hybrid-GAN method employs local gradient magnitude as a structure-aware weighting mechanism, enabling

adaptive blending that preserves anatomical details in edge-rich regions and semantic coherence in smooth tissue areas. Comprehensive evaluation on a balanced four-class Alzheimer's MRI dataset (12,491 images) with simulated 10-30% rectangular occlusions (missing-data imputation) demonstrated superior performance across multiple dimensions. The medical-specific GAN achieved the best reconstruction quality with MSE of 3679.84 ± 2474.27 (41.3% improvement over OpenCV, 48.7% over LaMa), ROI-based PSNR of 17.44 ± 19.05 dB, and ROI-based SSIM of 0.9796 ± 0.0128 (all PSNR/SSIM values are computed over the brain mask to avoid background-induced bias). The Hybrid-GAN achieved an MSE of 3808.94 ± 2243.87 ; the Attention-Guided GAN inference requires 8.49 ± 0.65 ms per image while the gradient-based blending step adds 0.48 ± 0.06 ms, yielding a Hybrid-GAN total inference time of approximately ≈ 8.97 ms per image. This total inference time is substantially faster than LaMa (≈ 23.24 ms; $\sim 2.6\times$ speedup) and comparable to the standalone GAN, while preserving diagnostic information. Critically, downstream classification experiments validated clinical utility: a VGG16 classifier achieved 96.05% accuracy on Hybrid-GAN inpainted images (only 0.96 percentage points below the original baseline of 97.01%), substantially outperforming OpenCV (94.72%), LaMa (94.77%), and Hybrid (OpenCV–LaMa) (95.20%). The Hybrid-GAN approach recovered 85.2% of diagnostic accuracy lost due to masking, with micro-averaged AUC of 0.9976, demonstrating effective preservation of disease-discriminative features across all severity categories. Future work includes extension to 3D volumetric inpainting for inter-slice consistency, multi-modal MRI integration leveraging complementary sequences (T1, T2, FLAIR), explicit evaluation on simulated k-space motion corruption and deblurring, adaptive mask-aware blending strategies using reinforcement learning, domain-specific fine-tuning on large scale medical datasets (UK Biobank, ADNI), prospective radiologist evaluation for clinical validation, and integration of uncertainty quantification and explainable AI techniques to enhance interpretability and regulatory readiness.

All the Declarations and Statements

Author Contributions Statement

Chhaya Yadav – Conceptualization, Methodology, and Software Implementation: Proposed the research idea, designed the two-stage GAN framework with attention gates, and implemented the hybrid inpainting model.

Sunita Yadav – Data Curation and Preprocessing: Collected and curated the Alzheimer's MRI dataset, performed preprocessing, augmentation, and managed data preparation for training and evaluation.

Arvind Panwar – Supervision, Formal Analysis, Writing – Review & Editing, and Project Administration: Supervised the overall research work, guided model development and evaluation, conducted critical analysis of results, and reviewed and edited the manuscript while managing project coordination.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Declaration

No Funding Received

Data Availability Statement

This study analyzed publicly available datasets. The results obtained and datasets can be found here: <https://www.kaggle.com/datasets/drsaeedmohsen/alzheimer-dataset>, accessed on: 2024-12-28

Ethical Declarations

The authors declare no involvement of human subjects and/or animals.

Acknowledgments

We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing

The authors used generative AI tools solely to improve the clarity, grammar, and readability of the manuscript, with all outputs carefully reviewed and validated under human supervision.

Abbreviations

The following abbreviations are used in this manuscript:

ADNI - Alzheimer's Disease Neuroimaging Initiative

AUC - Area Under the ROC Curve
CNN - Convolutional Neural Network
GAN - Generative Adversarial Network
GPU - Graphics Processing Unit
LaMa - Large Mask Inpainting
LR - Learning Rate
MD - Mild Demented
MOD - Moderate Demented
MRI - Magnetic Resonance Imaging
MSE - Mean Squared Error
NOD - Non-Demented
PSNR - Peak Signal-to-Noise Ratio
ROC - Receiver Operating Characteristic
ROI - Region of Interest
SMOTE-NM - Synthetic Minority Over-sampling Technique with Noise Modification
SSIM - Structural Similarity Index
VMD - Very Mild Demented

References

- [1] E. Twiss, C. McPherson, and D. F. Weaver, "Global diseases deserve global solutions: Alzheimer's disease," *Neurology International*, vol. 17, no. 6, p. 92, 2025.
- [2] W.-S. Tae, B.-J. Ham, S.-B. Pyun, and B.-J. Kim, "Current clinical applications of structural mri in neurological disorders," *Journal of Clinical Neurology (Seoul, Korea)*, vol. 21, no. 4, p. 277, 2025.
- [3] R. Obuchowicz, J. Lasek, M. Wodziński, A. Piórkowski, M. Strzelecki, and K. Nurzynska, "Artificial intelligence empowered radiology—current status and critical review," *Diagnostics*, vol. 15, no. 3, p. 282, 2025.
- [4] I. N. Sari, E. Horikawa, and W. Du, "Interactive image inpainting of large-scale missing region," *IEEE Access*, vol. 9, pp. 56 430–56 442, 2021.
- [5] Q. Wang, S. He, M. Su, and F. Zhao, "Image inpainting methods: A review of deep learning approaches," *Symmetry*, vol. 18, no. 1, p. 94, 2026.
- [6] R. G. Alexander, F. Yazdanie, S. Waite, Z. A. Chaudhry, S. Kolla, S. L. Macknik, and S. Martinez-Conde, "Visual illusions in radiology: untrue perceptions in medical images and their implications for diagnostic accuracy," *Frontiers in Neuroscience*, vol. 15, p. 629469, 2021.
- [7] J. C. Santos, H. Tomás Pereira Alexandre, M. Seoane Santos, and P. Henriques Abreu, "The role of deep learning in medical image inpainting: A systematic review," *ACM Transactions on Computing for Healthcare*, vol. 6, no. 3, pp. 1–24, 2025.
- [8] A. Telea, "An image inpainting technique based on the fast marching method," *Journal of graphics tools*, vol. 9, no. 1, pp. 23–34, 2004.
- [9] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1. IEEE, 2001, pp. I–I.
- [10] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, p. 24, 2009.
- [11] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [12] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2536–2544.
- [13] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Generative image inpainting with contextual attention," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5505–5514.
- [14] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 85–100.
- [15] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 4471–4480.
- [16] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky, "Resolution-robust large mask inpainting with fourier convolutions," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 2149–2159.
- [17] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "Repaint: Inpainting using denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 461–11 471.
- [18] Z. Wan, J. Zhang, D. Chen, and J. Liao, "High-fidelity pluralistic image completion with transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 4692–4701. Volume (), Issue 17
- [19] S. Iizuka, E. Simo-Serra, and H. Ishikawa, "Globally and locally consistent image completion," vol. 36, no. 4. ACM New York, NY, USA, 2017, pp. 1–14.
- [20] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "Edgeconnect: Structure guided image inpainting using edge prediction," in *Proceedings of the IEEE/CVF international conference on computer vision workshops*, 2019, pp. 0–0.
- [21] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning

- to leverage salient regions in medical images,” *Medical image analysis*, vol. 53, pp. 197–207, 2019.
- [22] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz et al., “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.
- [23] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European conference on computer vision*. Springer, 2016, pp. 694–711.
- [24] L. A. Gatys, A. S. Ecker, and M. Bethge, “Image style transfer using convolutional neural networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2414–2423.
- [25] K. Armanious, C. Jiang, M. Fischer, T. K\"ustner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, “Medgan: Medical image translation using gans,” *Computerized Medical Imaging and Graphics*, vol. 79, p. 101684, 2020.
- [26] W. H. Pinaya, P.-D. Tudosiu, J. Dafflon, P. F. Da Costa, V. Fernandez, P. Nachev, S. Ourselin, and M. J. Cardoso, “Brain imaging generation with latent diffusion models,” *arXiv preprint arXiv:2209.07162*, 2022.
- [27] M. “Ozbey, O. Dalmaz, S. U. Dar, H. A. Bedel, S. . “Ozturk, A. G“ung“or, and T. C. ukur, “Unsupervised medical image translation with adversarial diffusion models,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 12, pp. 3524–3539, 2023.
- [28] D. Wang, Y. Kang, Y. Chen, Y. Gao, and S. Xu, “Sain: structure-aware image inpainting for large missing areas,” *Journal of King Saud University Computer and Information Sciences*, 2026.
- [29] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, “Image inpainting: A review,” *Neural Processing Letters*, vol. 51, pp. 2007–2028, 2020.
- [30] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof, “Diffusion models for medical image analysis: A comprehensive survey,” *arXiv preprint arXiv:2211.07804*, 2022.
- [31] K. Velu and N. Jaisankar, “Design of a cnn–swin transformer model for alzheimer’s disease prediction using mri images,” *IEEE Access*, 2025.
- [32] Kaggle, “Alzheimer’s mri preprocessed dataset,” <https://www.kaggle.com/datasets/drsaeedmohsen/alzheimer-dataset>, accessed: 2024-12-28.
- [33] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [34] L. Zhou, A. Jain, A. K. Dubey, S. K. Singh, N. Gupta, A. Panwar, S. Kumar, T. A. Althaqafi, V. Arya, W. Alhalabi et al., “Fpa-based weighted average ensemble of deep learning models for classification of lung cancer using ct scan images,” *Scientific Reports*, vol. 15, no. 1, p. 19369, 2025.

Authors’ Profiles



Chhaya Yadav is a PhD research scholar in the Department of Computer Science and Engineering at Galgotias University, India. Her research focuses on advancements in deep learning techniques for high-fidelity image inpainting. Her broader research interests include artificial intelligence, deep learning, and computer vision, with publications in international conferences.



Dr. Sunita Yadav is a distinguished academic leader and Professor in Computer Science and Engineering with over two decades of experience in teaching, research, and institutional administration. She is currently serving as Dean (Computer Science and Allied Branches) at Inderprastha Engineering College, Ghaziabad, Uttar Pradesh, India. Her research interests include artificial intelligence, deep learning, information retrieval, and computer networks, with notable contributions in medical image analysis, particularly in cataract detection using deep learning. Dr. Yadav has published over 40 research papers in reputed SCI/Scopus-indexed journals and conferences and has led funded research projects as Principal Investigator. She has supervised Ph.D. and postgraduate scholars and has played a significant role in accreditation processes such as NAAC and NBA. She is also actively involved in fostering industry-academia collaborations and promoting student-centric, outcome-based education.



Dr. Arvind Panwar is a dedicated academician with more than a decade of experience teaching computer science and engineering courses at the undergraduate and postgraduate levels. He is currently serving as an Assistant Professor at Galgotias University since January 2023. Before that, he taught at reputed institutions including Dr. Akhilesh Das Gupta Institute of Technology & Management and Northern India Engineering College. He holds a Ph.D. in Computer Science from Guru Gobind Singh Indraprastha University, where his research focused on blockchain-based health record management systems. He also earned his M.Tech. and B.Tech. degrees in Computer Science & Engineering from the same university. Dr. Panwar has an extensive research portfolio, with 14 international journal articles, 20 conference papers, and 20 book chapters to his credit. His research interests include blockchain, cryptography, network security, data mining, big data, and cloud computing. He has been granted 8 patents and several more are under review. He is currently editing three books: *Data Analytics and Artificial Intelligence for Predictive Maintenance in Industry 4.0* (Bentham), *Qubits Unveiled: Quantum Computing Solutions for Efficient Supply Logistics* (Nova Publications), and *Energy Efficient Internet of Things-Based Wireless Sensor Network* (Wiley-Scrivener Publishing LLC), reflecting his expertise in modern technologies and industrial applications.

How to cite this paper:

Chhaya Yadav, Sunita Yadav, Arvind Panwar, "Two-stage GAN with Attention Gates for Brain MRI Inpainting: A Hybrid Framework for Preserving Diagnostic Features in Alzheimer's Disease Classification", International Journal of Information Technology and Computer Science(IJITCS), Vol.18, No.5, pp.171-189, 2026. DOI:10.5815/ijitcs.2026.05.11