

# Self-supervised Diffusion-based 3D Brain MRI Enhancement for Anomaly Detection and Segmentation

Ram Singh<sup>1</sup>, Arshdeep Singh Bhankhar<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Punjabi University, Patiala-147002, Punjab, India

<sup>2</sup>Department of Computer Science and Engineering, Thapar Institute of Engineering & Technology, Patiala-147001, Punjab, India

## ARTICLE INFO

Received: 01 Nov 2024

Accepted: 17 Dec 2024

Published: 28 Dec 2024

## ABSTRACT

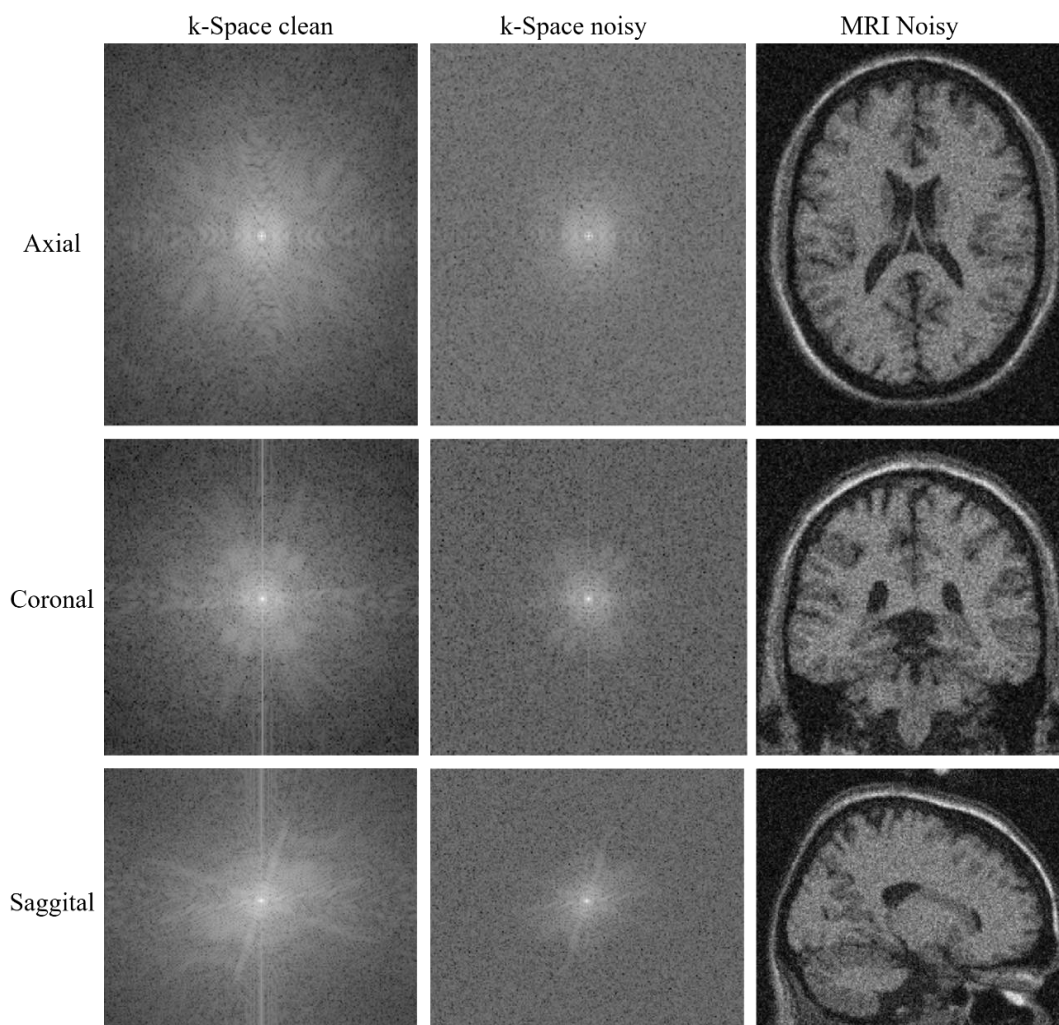
Magnetic resonance imaging (MRI) is a widely used non-invasive imaging technique for brain image acquisition in medical diagnostic analysis and treatment planning. During acquisition, MRI data are corrupted by various factors, such as random Gaussian, Rician, and Rayleigh noise probabilities, signal distortions, and motion artifacts inherent to the MRI system. Denoising is a major pre-processing imaging task in computer-aided diagnostic (CAD) systems applied to suppress the noise effect and regenerate a high-quality enhanced image with maximum useful image information. Traditional filtering-based MRI denoising and restoration methods demonstrate limited-quality image generation with single-image denoising techniques, whereas these methods fail to preserve essential information, such as edge features, in the restored images with large input datasets. Deep-learning-based (DL) image denoising methods have demonstrated significant performance in natural image restoration applications. DL methods have great potential to preserve and enhance the image quality of reconstructed images without losing important image information. The increasing complexity of medical image analysis techniques underscores the need for robust and advanced anomaly detection and automatic medical image feature classification. Existing methods have certain limitations and face challenges in identifying and capturing the desired information patterns of anomalies and often limit their usage to specific types of lesions in brain MRI. This study addresses the challenge of using a DL-based novel unsupervised learning ‘reverse-encoder-decoder’ network model to generate different types of MRI pathologies. The proposed unsupervised anomaly detection methods were evaluated using BraTS2021, IXI, and simulated BrainWeb MRI datasets and demonstrated improved anomaly detection, image restoration accuracy and segmentation performance in terms of low error and higher peak signal-to-noise ratio (PSNR), structural similarity (SSIM), dice score metrics values.

**Keywords:** brain MRI, enhancement, anomaly detection, unsupervised learning, diffusion learning, denoising encoders-decoders, segmentation.

## 1. INTRODUCTION

Analysing brain MR images for anomaly detection is a major challenging task in medical image processing. The detection of brain anomalies, such as brain tumours, multiple sclerosis, stroke regions, and neurodegenerative disease identification, and accurate detection are vital in clinical diagnosis. An experienced radiologist may visually inspect MRIs to analyse structural abnormalities, asymmetries, and unusual signal patterns in the acquired images; however, visual MRI analysis can provide limited or missing relevant pathological information in 5–10% of images [1]. Moreover, when the quality of the acquired MR images degrades, anomaly detection becomes increasingly challenging. Magnetic resonance imaging (MRI) is one of the preferred imaging modalities used in clinical diagnosis and treatment decisions, disease progress monitoring, and analysis. However, this manual procedure is subjective, time-consuming, and has limited reliability, with the possibility of prone human errors, especially in subtle or early-stage mild abnormalities. Advances in medical imaging techniques have led to exponential growth and an increase in medical imagery volume. The increasing complexity and volume of data necessitate more sophisticated and robust medical image analysis methods [2]. Traditional image anomaly detection methods are based on conventional supervised manual parameter-based techniques designed to identify specific image-based pathologies such as brain tumor detection [3], brain stroke detection [4], and white matter (WM) hyperintensity correction [5]. Despite their effectiveness in single images and specific scenarios, these methods inherently suffer from inhomogeneous intensity biases that contribute to additional artifacts in the recovered images. The limitations of these methods demonstrate

the major implications that narrow the scope of their effectiveness in detecting the desired pathologies. Figure 1 shows clean noise free images, corresponding noise and bias-field affected MRI images of each spatial orientation's axial, coronal and sagittal view planes.



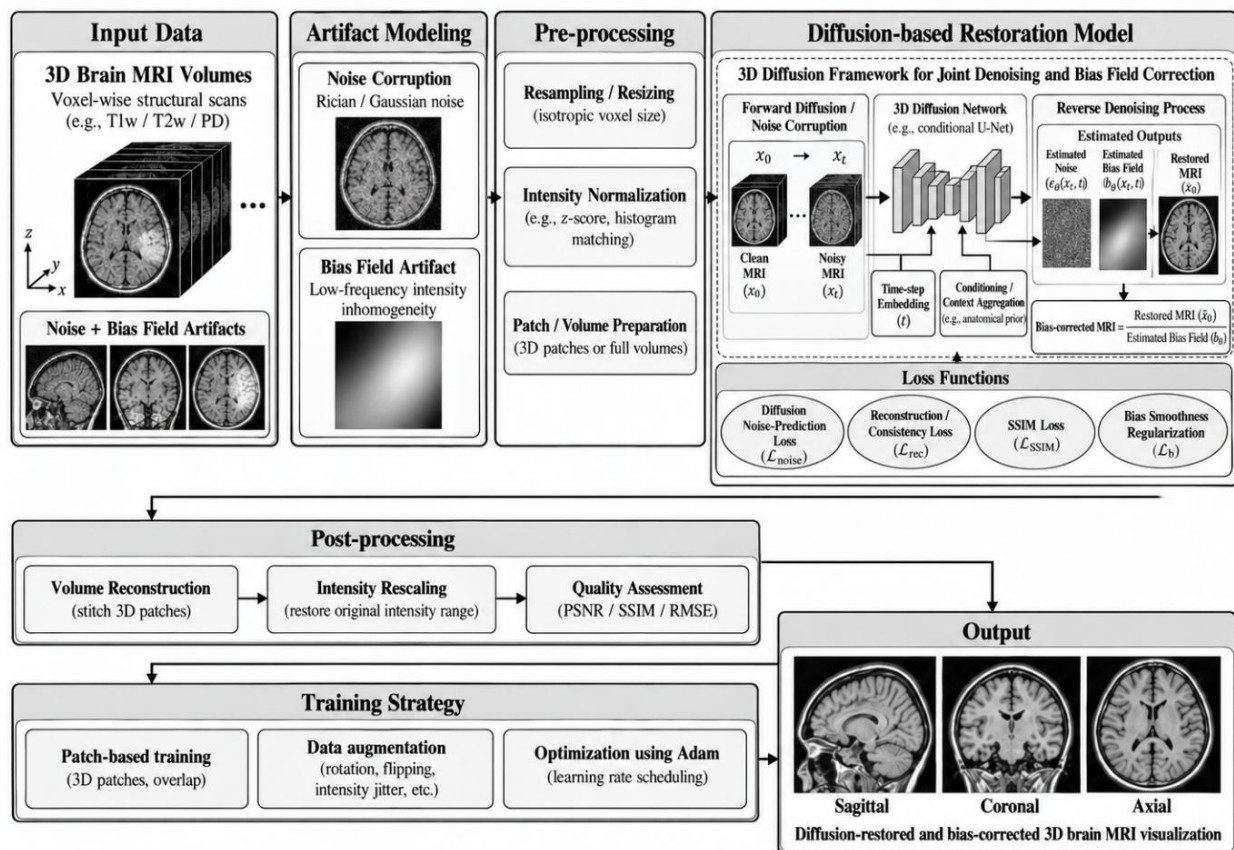
**Fig.1 Complex domain acquired k-space MRI and corresponding observed noise affected images in each orientation views**

## 2. RELATED WORK

In the recent past, significant developments have been made in the support of techniques and tools for the assistance of radiologists. DL-based autoencoders have been widely used for image denoising and restoration [6]. Convolutional neural network (CNN) models have proven successful in many medical imaging tasks, producing enhanced medical images while effectively preserving important features [7]. Supervised machine learning (ML) anomaly detection methods require large, labelled, and heterogeneous training datasets that are normally not of good quality. Moreover, annotation is costly and difficult to acquire [8]. Unsupervised DL-based anomaly detection methods have emerged as promising alternatives for detecting anomalies in acquired brain images. These methods require a healthy cohort of patients for DL network model training, after which the models can be applied to the detection of abnormal image region patterns. However, the major challenge with unsupervised methods is the utilization of a singular or limited number of conditions in self-supervised or weakly supervised feature learning methods.

Autoencoders and variational autoencoders have demonstrated reliable performance in training and fast inference, producing blurred images for unsupervised anomaly detection [6]. Furthermore, various methods have been proposed to better understand the image context using latent image representation and to add spatial latent

information [9], multiresolution skip connections with dropout [10], and denoising as regularization. Other approaches include image restoration [11], uncertainty estimation for anomaly segmentation [12], and vector-quantization-based variational autoencoders [13]. Autoencoder models are typically trained on healthy input data by minimizing the reconstruction mean squared error (MSE) closer to zero as the main optimization objective. With careful normalization and morphological post-processing techniques, better anomaly detection performance can be achieved [14]. GAN models have unstable training problems and suffer from mode collapse because they fail to retain the fine details of anatomical coherence. To overcome this problem, an inpainting generator model was proposed to inpaint the missing information in affected image regions [15]. Existing approaches aim to recover all image details simultaneously to preserve the structural details in the restored image. The proposed patch-based diffusion approach restores images by inpainting missing information in the generic image domain. The patch-based diffusion method adequately reconstructs brain MRI, restoring the global contextual information of individual brain structures by estimating the individual image patches. Some existing patch-based reconstruction methods either neglect neighboring patch information [16] or restore image patches from high-level noise-affected images, which also affects neighboring image patches [17]. Considering the limitations of existing patch-based approaches, it is of interest to develop a patch-based diffusion model for image restoration by denoising[18]. Figure 2 shows the architecture of the diffusion-based MRI noise removal and bias field correction for intensity inhomogeneity removal.



**Fig. 2 Architecture of the proposed diffusion learning MRI denoising and bias field correction in 3D brain images**

### 3. PROPOSED METHOD

The patch-based diffusion model uses a real input image  $f$  with width  $W$ , depth  $D$  and channels  $C$  corrupted by additive white random Gaussian noise. The input image is decomposed into overlapping small patches of  $c \times m \times n$

dimensions, and patches  $\mathcal{P}_n \in \mathbb{R}^{C \times H \times W}$  with  $h < H, w < W$  and  $n = [1, 2, 3, \dots, N]$ . Each small  $n$ th patch was reconstructed to estimate the maximum brain anatomy, such that the contextual brain information was restored from the unperturbed surrounding information. The motivation for the same is to preserve essential minute details and restore better anatomical coherence in each individual patch. To avoid the test time inception of structural anomalies anywhere in the brain regions, the patch-based denoising and information restoration approach promisingly contributes to the reconstruction of quality images. A network of sliding window denoising kernels is applied to add and remove noise from each patch, which is evenly distributed across the entire image space. After performing the denoising operation on all image patches of each orientation dimension, all individual patches were stitched to reconstruct a complete image. This strategy allows the processing of every region of an image, considering the spatial contextual information of each patch and its surroundings, which is expected to be particularly helpful for detecting and recovering anomalies in image patches.

### 3.1 Denoising diffusion model

The diffusion model is first applied to image patches for gradual degradation by adding random noise and subsequently applying a reverse denoising process to learn the image features. First, during the forward process, random Gaussian noise  $\eta \sim \mathcal{N}(0, \mathbf{I})$  is added to the image  $f_0$  with a predefined schedule  $t_1, \dots, t_T$ , which is defined as follows:

$$f_0 \sim q(f_t | f_0) = \mathcal{N}(\sqrt{\alpha_t} f_0, (1 - \alpha_t) \mathbf{I}), \text{ with } \alpha_t = \prod_{s=0}^t (1 - \bar{\alpha}_s) \quad (1)$$

where the time-interval step  $t$  is captured from an evenly distributed period  $t \sim \text{normal}(1, 2, 3, \dots, T)$ , and controls the noise added to the image  $f_0$ . When the time  $t = T$ , the input image is converted into a pure Gaussian distributed noise, that is  $f_t = \eta \sim \mathcal{N}(0, \mathbf{I})$ , when  $t = 0$ ,  $f_t$  becomes  $f_0$ .

In reverse processing, the target is to recover the degradation of the forward processing to map the original image  $f_0$  as defined below:

$$f_0 \sim \mathcal{P}_\theta(f_t) \prod_{t=1}^T \mathcal{P}_\theta(f_{t-1} | f_t), \text{ with } \mathcal{P}_\theta(f_{t-1} | f_t) = \mathcal{N}(\mu_\theta(f_t, t), \sum_\theta(f_t, t)) \quad (2)$$

where  $\mu_\theta$  and  $\sum_\theta(\cdot)$  are approximated by a neural network using parameter  $\theta$ . According to the strategy of [19] and the U-Net model [20], the above task is accomplished by keeping  $\sum_\theta(f_t, t) = \frac{1-\alpha_{t-1}}{1-\alpha_t} \beta_t \mathbf{I}$  fixed. For tractable loss calculations, the variational lower bound (VLB) was used. The Bayes rule probabilistic formulation is applied by conditioning  $q(f_{t-1} | f_t)$  on  $f_0$ , and minimizing the VLB to approximate the simpler loss derivation by  $\mathcal{L}_{\text{Simple}} = \|\eta - \eta_0(f_t, t)\|^2$ . The  $L_1$  - error loss is used to estimate directly  $f_0^{\text{Rec}} \sim \mathcal{P}_0(f_0 | f_t, t)$ , which leads to  $\mathcal{L}_{\text{Rec}} = |f_0 - f_0^{\text{Rec}}|$ . To capture images using diffusion models, a stepwise strategy was applied to denoising for all time intervals beginning from  $t = T$ . It comes with time-consuming computations and fixed time steps to estimate the reconstruction as  $f_0^{\text{Rec}} \sim \mathcal{P}_0(f_0 | f_t, t)$  with fixed interval  $t_{\text{test}}$ .

### 3.2 Realization of Patch-based diffusion model

In the forward image processing and inverse reconstruction, small image patches of  $m \times n$  dimensions were created as mentioned above. The training phase begins by sampling the image patches from fixed indexed sub-image locations created from the input image  $f$  into a  $K$  numbers of image patches evenly spaced over the spatial image space. The total number of patches  $K$  derived from the input image  $f$  as:

$$K = \left\lceil \frac{W-w}{w} \right\rceil + \left\lceil \frac{H-h}{h} \right\rceil + 2 \quad (3)$$

where  $\lceil \cdot \rceil$  is the ceiling function. An index was created from the entire image space an index  $k$  is created. During forward processing of diffusion, noise-degraded images  $f_t$  were sampled at each patch-indexed location  $\mathcal{P}_k$ . Under a binary mask  $\text{Mask}_{\mathcal{P}} \in \mathbb{R}^{CHW}$ , the image pixels similar to the patch  $\mathcal{P}_k$  were set to one, and the remaining pixels that did not match  $\mathcal{P}_k$  were updated to zero. The elementwise multiplication operation between the selected image patch and mask operator provides a partially noise-affected image, as defined below:

$$\bar{f}_t = f_t \odot \text{Mask}_p + f_0 \odot \neg \text{Mask}_p \quad (4)$$

The evaluation process began by considering each indexed patch  $\mathcal{P}_k$  for every patch  $k \in [0, \dots, K]$  and subsequently performing diffusion operations on each patch  $\mathcal{P}_k$ .

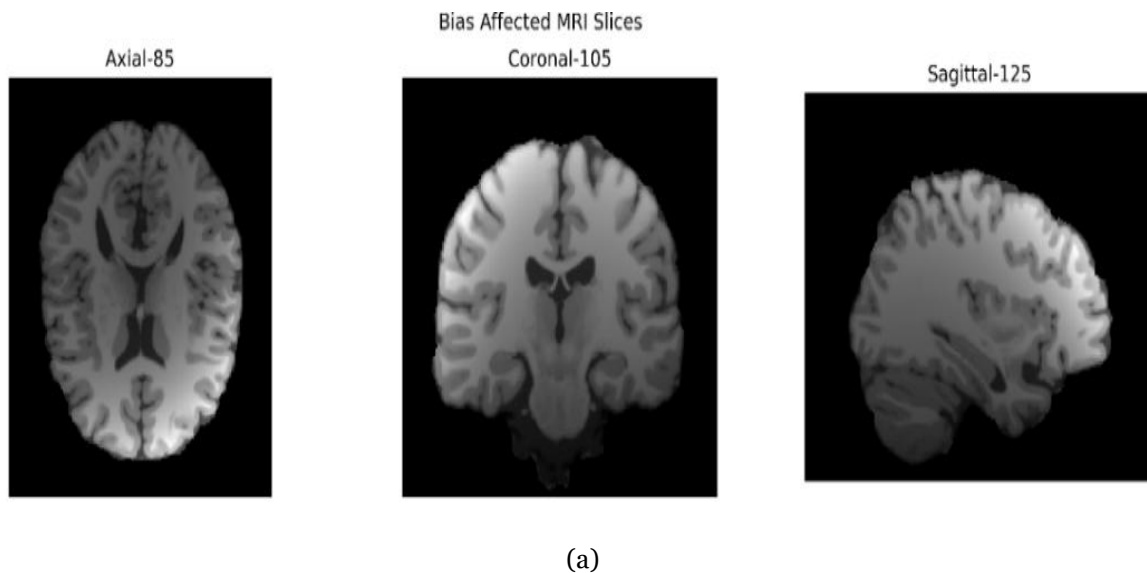
### 3.3 Anomaly Detection

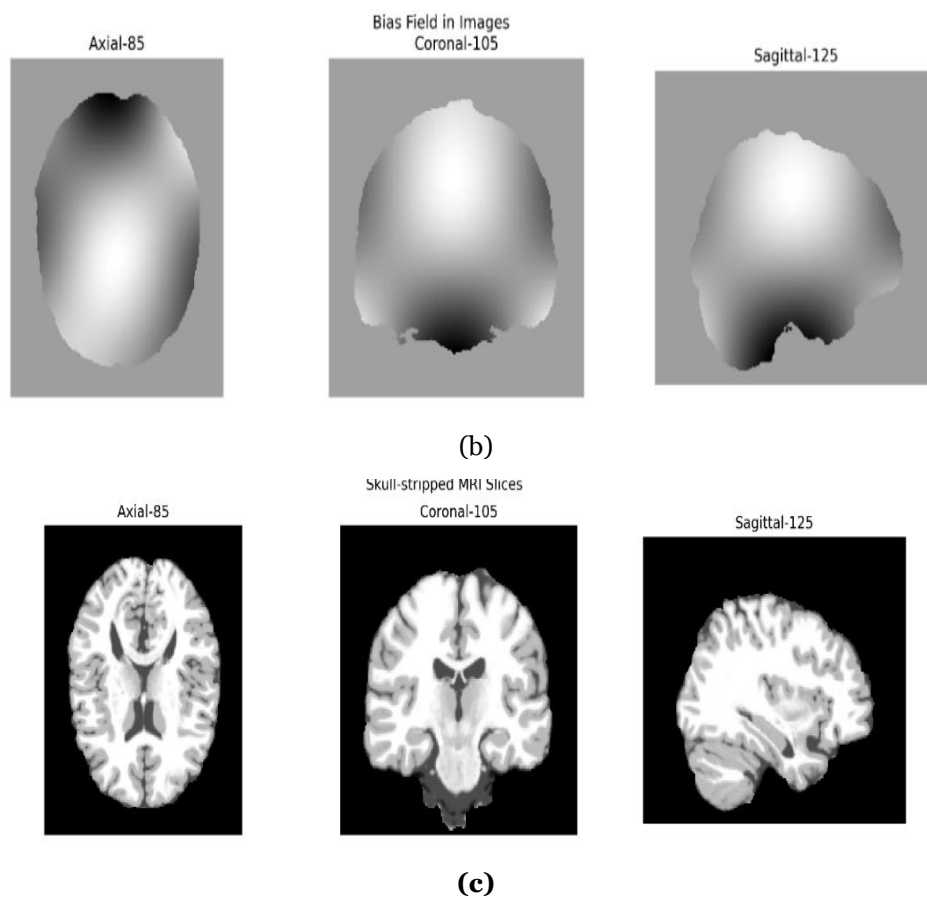
During reverse image reconstruction, patch-level anomaly detection is performed by quantifying the information deviation between the queried and corresponding reconstructed patch. All image-region-based components required for restoration in the reconstructed images for anomaly detection and correction were incorporated during anomaly quantification. The training phase preserves the region-specific features in the filtering kernels of the neural network for restoration of the corresponding image patches during the decoding or reverse reconstruction process of the model. During training, anomaly quantification follows the difference between the image patch pairs to estimate the loss for encoder training by minimizing the mean squared error of the input and reconstructed samples.

## 4 EXPERIMENTAL EVALUATION

### 4.1 Dataset

The performance of the proposed diffusion model was evaluated using brain MR image results by establishing multiple baselines for diffusion-based anomaly detection on BraTS2021, IXI T1w and the simulated BrainWeb T1w MRI dataset. This is a publicly available and widely used open-source MRI dataset for experimentation in medical imaging research. We included image paired T1w and T2-w brain MRI from each dataset, considering 360 from each dataset. Therefore, 600 samples were used for training and 120, considering 30 images from each modality for testing. For validation, random image samples were used from both the MRI datasets. Skull stripping was also applied using HD-BET, and the black borders were subsequently removed from the fixed MRI slices of dimensions  $192 \times 192 \times 160$ . Intensity inhomogeneities were also corrected before using the sample using the standard available tool N4ITK [21]. Figure 3 shows the skull-stripped bias magnetic field affected images of each orientation of an MRI volume. and corrected images.





**Fig. 3** Top row (a) contains skull-stripped bias field-affected images; (b) the middle row shows the extracted bias field; and (c) the last row shows the bias field-corrected images.

To accelerate the execution and save resources, the volume resolution was reduced to  $96 \times 96 \times 80$  voxels by removing 15 top and bottom MRI slices parallel to the transverse spatial z-axis plane. The details of the simulation settings, parameter configurations and datasets used for evaluating the proposed method. It also presents the results and a performance comparison. The implemented approach runs on Python using an Intel(R) Core (TM) i7-10700 CPU, an NVIDIA GeForce RTX 3070 GPU, and 16 GB of memory. Additionally, the parameter configurations for the proposed method are outlined in Table 1.

**Table 1: Parameter configuration of proposed method**

| Parameter's No | Parameter's Name          | Values       |
|----------------|---------------------------|--------------|
| 1.             | Filters count             | 64, 128, 256 |
| 2.             | Kernel size               | 3*3          |
| 3.             | Dropout size              | 0.5          |
| 4.             | Learning size             | 0.0001       |
| 5.             | Epochs count              | 100          |
| 6.             | Batch size                | 16           |
| 7.             | Optimizer                 | Adam         |
| 8.             | Loss                      | MSE          |
| 9.             | Noise level – step size 2 | 1%-15%       |

These comparisons included autoencoders, variational autoencoders [22] [8], and sequential variational encoders [23]. Simple threshold and GAN-based methods were compared [24], [25]. All baseline parameters were set at par with the original values provided in the respective publications. For example, the values for the variational autoencoder and sequential autoencoder were set to  $\beta_{VAE} = 0.001$ , the latent space to 128 and the learning rate to  $1e - 4$ . The noise type is added as a random Gaussian probability, which is known to affect the MR signal frequencies during acquisition in a complex domain. For model training, uniformly sampled time intervals  $t \in [1, T]$  with maximum  $T = 1000$  were used, and for testing time, a fixed value of 500 was used as  $t_{test} = \frac{T}{2}$ . A linear schedule was used as  $\beta_t$ , between  $1e - 4$  to  $2e - 2$ , and the U-Net architecture was used for image denoising [26]. For each dimension of the input channel  $C_f \in [128, 128, 256]$ , the U-Net model consisted of three residual layers with down sampled convolutions. This architecture is mirrored in the upsampling path with transposed convolution. A skip connection was used to connect the outputs of each layer resolution. The residual of each layered block was batch-normalized using sigmoid linear activation [27] before the convolution. All models were trained using a maximum of 1600 epochs, and the best model checkpoints were used to determine the performance of the clean images for testing purposes. All input data were processed on a slice-by-slice basis with uniform slice sampling and replacement during training. Iteratively, all slices were used to reconstruct the full volume during the testing. The proposed models are trained online on <https://colab.google/> platform. Adam optimizer is used with learning rate  $1e - 5$  and batch size of 32.

#### 4.2 Anomaly Scoring

The aim of training the models was to minimize the  $l1 - norm$  error between the input and its corresponding reconstructed image. The pixel-wise reconstruction error was calculated to measure the anomaly score as  $\Delta_{AS} = |f_0 - f_0^{Rec}|$ , where high values of  $\Delta_{AS}$  indicate large reconstruction errors and dissimilarities between the reconstructed and the target images. Different dimensions of median filtering kernels of size 3-5 were used, and were also applied to minimize the reconstruction error along with brain mask erosion for three iterations before binarization of the anomaly score  $\Delta_{AS}$ . With the binarization of  $\Delta_{AS}$ , connected component analysis was performed to remove segmentations with  $< 7$  voxels. An iterative Dice score was calculated based on a greedy search for different threshold values for binarization  $\Delta_{AS}$ , using a noise-affected validation set. The best threshold values were used to compute the average Dice score using the degraded test set. The average area under the precision-recall curve and mean absolute reconstruction error ( $l1$ ) of the clean dataset are given.

#### 4.3 Statistical Analysis

Statistical analysis was conducted to test the performance by employing permutation testing using the MLXtend Python library package [28]. The test scores were calculated using the two models' average difference in the Dice values for each permutation. The  $p - value$  result was determined by calculating the number of times the average difference was equal to or greater than the sample differences divided by the total number of permutations. Table 2 and 3 lists the Dice score and denoising results with comparison between other existing MRI denoising and reconstruction methods for anomaly detection.

**Table 2: Comparison of Dice score of with other state-of-the-art MRI reconstruction methods**

| Models            | BrainWeb   |
|-------------------|------------|
|                   | Dice (%)   |
| Threshold         | 19.69      |
| Autoencoders (AE) | 32.87±1.25 |
| Variational AE    | 31.11±1.50 |
| Sequential AE     | 33.32±0.14 |
| Denoising AE      | 37.05±1.42 |

|                                |            |
|--------------------------------|------------|
| Fast-AnoGAN                    | 24.16±2.94 |
| DDPM                           | 40.67±1.21 |
| pDDPM+random sampling+ LRec    | 44.47±2.34 |
| pDDPM+fixed sampling+ LRec     | 47.81±1.15 |
| pDDPM+fixed sampling+ l1-error | 48.25±0.87 |

**Table 3: Comparison of MRI denoising results with other MRI datasets**

| Datasets     | PSNR (dB) values corresponding to different noise levels |       |       |       |       |       |       |
|--------------|----------------------------------------------------------|-------|-------|-------|-------|-------|-------|
|              | 3%                                                       | 5%    | 7%    | 9%    | 11%   | 13%   | 15%   |
| BraTs2021    | 36.35                                                    | 35.67 | 34.87 | 34.03 | 33.24 | 32.55 | 31.93 |
| IXI T1w      | 38.51                                                    | 37.95 | 37.27 | 36.56 | 35.78 | 35.09 | 34.44 |
| BrainWeb T1w | 39.55                                                    | 38.73 | 38.11 | 37.55 | 37.01 | 36.49 | 36.04 |
|              | SSIM values corresponding to different noise levels      |       |       |       |       |       |       |
|              | 3%                                                       | 5%    | 7%    | 9%    | 11%   | 13%   | 15%   |
| BraTs2021    | 0.944                                                    | 0.933 | 0.919 | 0.902 | 0.885 | 0.868 | 0.852 |
| IXI T1w      | 0.970                                                    | 0.966 | 0.961 | 0.955 | 0.948 | 0.942 | 0.935 |
| BrainWeb T1w | 0.827                                                    | 0.821 | 0.818 | 0.815 | 0.815 | 0.813 | 0.812 |

Table 4 provides the performance analysis across different noise levels (1%-15%) for BrainWeb, IBSR, and BT datasets shows that PSNR, SSIM, and MSE values degrade as noise increases. At 1% noise, BrainWeb achieves the best performance with PSNR of 44.26, SSIM of 0.981, and MSE of 0.032281, while IBSR and BT have slightly lower values. As noise increases to 15%, PSNR drops significantly to 32.02 (BrainWeb), 34.81 (BraTS2021), and 28.94 (IXI), and SSIM declines to 0.825, 0.775, and 0.771, respectively. MSE increases with noise, reaching 0.57784 (BrainWeb), 0.548236 (BraTS2021), and 0.48868 (IXI), confirming that higher noise levels reduce image quality and segmentation performance.

**Table 4: Image enhancement results on three MRI datasets by denoising in terms of PSNR, SSIM and MSE**

| Noise levels (%) | BraTS2021 |       |          | IXI T1w |       |          | BrainWeb T1w |       |          |
|------------------|-----------|-------|----------|---------|-------|----------|--------------|-------|----------|
|                  | PSNR      | SSIM  | MSE      | PSNR    | SSIM  | MSE      | PSNR         | SSIM  | MSE      |
| 1                | 44.26     | 0.981 | 0.032281 | 43.32   | 0.971 | 0.047996 | 42.3         | 0.956 | 0.055125 |
| 2                | 43.52     | 0.967 | 0.083101 | 42.26   | 0.959 | 0.084994 | 41.34        | 0.957 | 0.067745 |
| 3                | 42.18     | 0.97  | 0.067822 | 41.17   | 0.945 | 0.075404 | 40.01        | 0.929 | 0.147986 |
| 4                | 42.31     | 0.946 | 0.191754 | 41.71   | 0.947 | 0.184923 | 39.7         | 0.919 | 0.108688 |
| 5                | 40.05     | 0.955 | 0.186116 | 39.42   | 0.937 | 0.16689  | 38.61        | 0.922 | 0.24988  |
| 6                | 39.07     | 0.951 | 0.187156 | 40.79   | 0.921 | 0.247623 | 39.15        | 0.916 | 0.254011 |
| 7                | 38.85     | 0.923 | 0.286237 | 39.55   | 0.891 | 0.249347 | 37.07        | 0.933 | 0.263658 |

|    |       |       |          |       |       |          |       |       |          |
|----|-------|-------|----------|-------|-------|----------|-------|-------|----------|
| 8  | 39.41 | 0.906 | 0.26044  | 36.92 | 0.87  | 0.260648 | 36.81 | 0.922 | 0.330221 |
| 9  | 38.38 | 0.889 | 0.279678 | 36.00 | 0.845 | 0.468686 | 34.77 | 0.902 | 0.314649 |
| 10 | 38.28 | 0.874 | 0.264685 | 35.67 | 0.884 | 0.27521  | 35.95 | 0.834 | 0.519307 |
| 11 | 33.82 | 0.871 | 0.286217 | 33.46 | 0.866 | 0.258345 | 36.97 | 0.902 | 0.332965 |
| 12 | 33.23 | 0.826 | 0.551359 | 36.35 | 0.831 | 0.58898  | 35.02 | 0.849 | 0.284949 |
| 13 | 33.61 | 0.893 | 0.62387  | 34.99 | 0.885 | 0.393352 | 30.42 | 0.89  | 0.322952 |
| 14 | 31.00 | 0.863 | 0.726263 | 32.56 | 0.873 | 0.557915 | 29.56 | 0.842 | 0.363273 |
| 15 | 32.02 | 0.825 | 0.57784  | 34.81 | 0.775 | 0.548236 | 28.94 | 0.771 | 0.48868  |

For patch-based diffusion, kernel dimensions were set to  $m = n = \frac{H}{2} = \frac{W}{2} = 48$ . Similar to denoising autoencoders, the diffusion model shows relatively high-quality image restoration performance on the BrainWeb dataset, with significant statistical scores of Dice with probability ( $p < 0.05$ ). The quality-wise patch-based diffusion model produced better images and smaller reconstruction errors than the normal denoising diffusion probabilistic model. The input patch-size  $60 \times 60$  pixel image provided the best performance. Furthermore, the peak denoising and reconstruction performance occurred when the noise level was at the test time  $t_{test} = 400$ . Figure 4 shows the segmented images after performing denoising with path-based diffusion processing on Brainweb skull-stripped T1w 3D MRI volume. Figure 4 show the architecture of the proposed model for

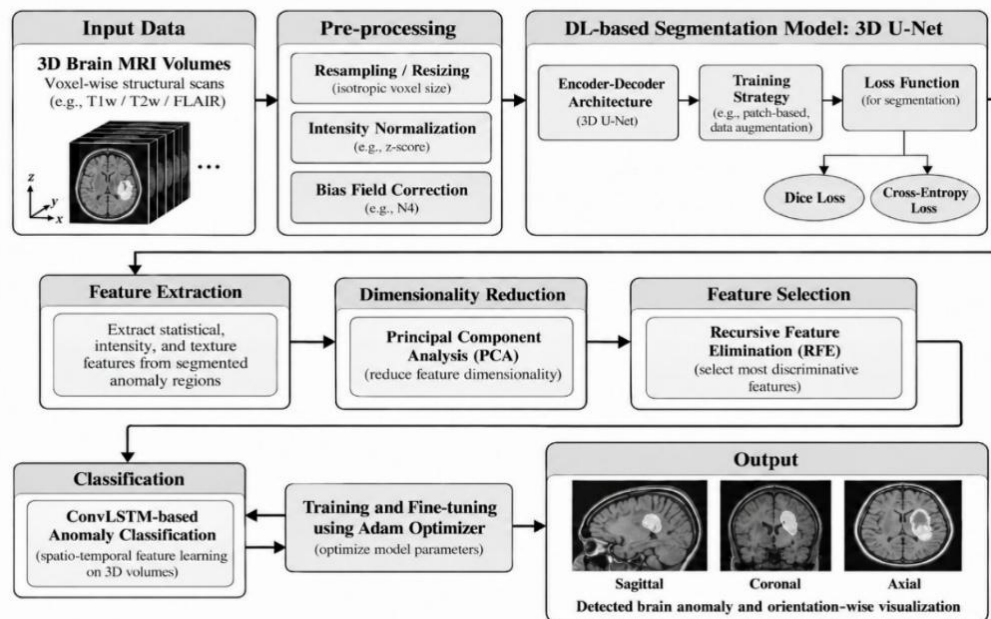
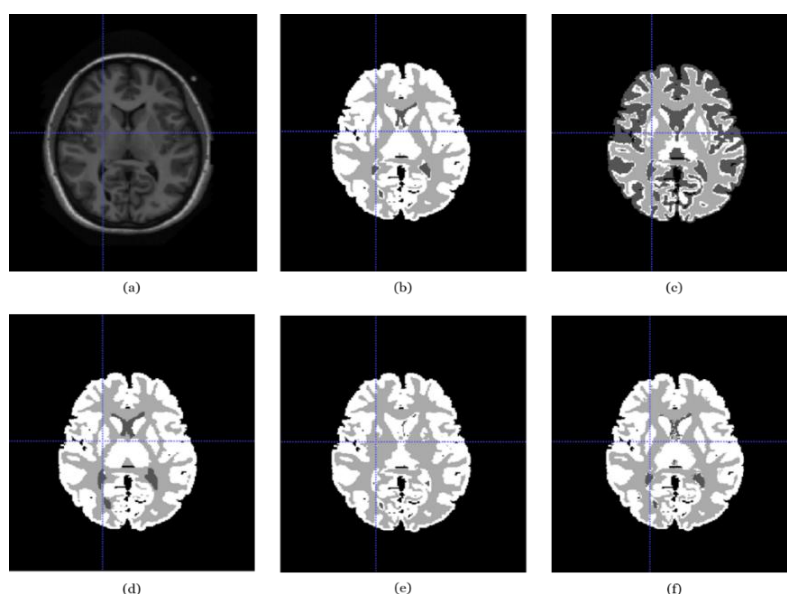


Fig. 4 The proposed architecture model of tissue segmentation in the denoised and bias field corrected brain MRI.

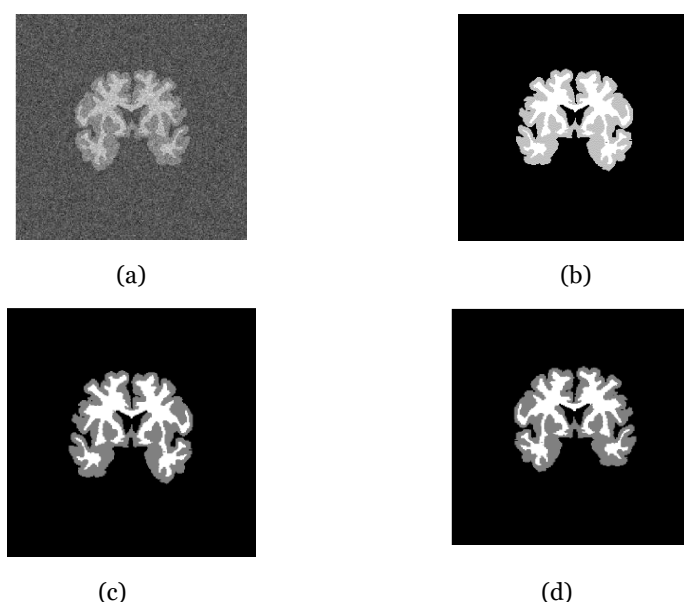
#### 4.4 Segmentation Analysis

The performance evaluation on the brain tumour BraTS2021 dataset shows that segmentation with bias-removed images yields superior results compared to segmentation with biased images as shown in Figure 8. The PSNR for bias-removed images is 42.245, significantly higher than 31.469 for biased images, indicating better noise reduction. The SSIM is also improved at 99.2517% versus 93.5035%, demonstrating enhanced structural similarity. Additionally, the Dice coefficient is 99.9565% for bias-removed images, compared to 98.8659% for biased images, reflecting more accurate segmentation performance.

Figure 5 shows the segmentation performance of the proposed method on each BrainWeb T1w dataset that demonstrated improved segmented results align with ground truth structures in MRI images. The proposed method attained accuracy of 99.26%, 99.99%, 99.91% for BraTS2021, IXI, and BrainWeb datasets respectively. Figure 5(a) through (f), illustrates the segmented slices with proposed method that attains lower error rates of 0.0015 on skull-stripped BrainWeb, 0.0137, on BraTS2021 0.00095 reflecting the improved classification of soft tissue regions in the respective image space segmented pixels or regions. Lower error rates signify more precise identification of brain structures and abnormalities.

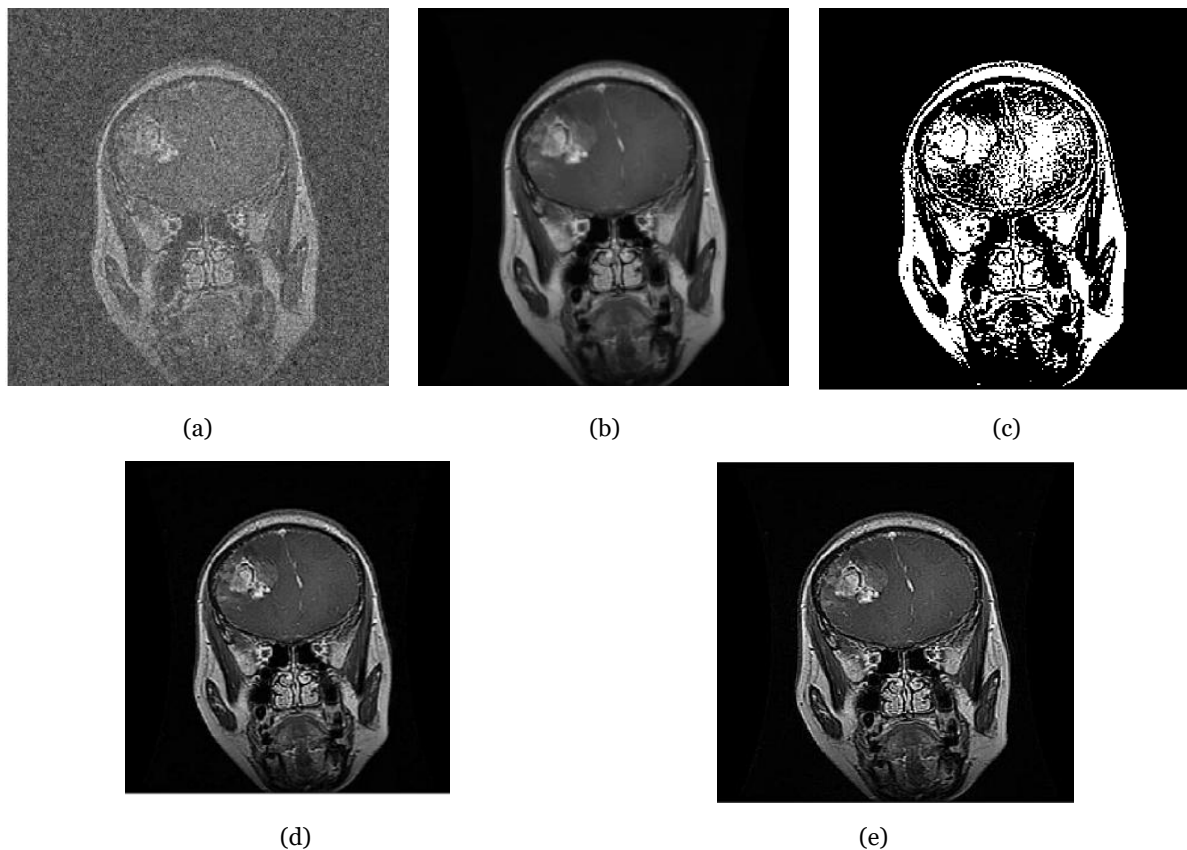


**Fig. 5 (a) Denoised enhanced image by noise removal (b) through (f) segmentation of soft-tissue regions setting different levels of intensity thresholding**



**Fig 6: Visualization of (a) Noise degraded skull-stripped BrainWeb T1w coronal slice (a) ground-truth image (c) and (d) segmented images**

Figure 6 show the noise degraded slices of BraTS2021 T1w dataset and corresponding denoising and segmented slices detecting tumour regions. Figure 7, Figure 8 and Figure 9



**Fig 7: Visualization of (a) noise degraded by 9% noise level of BraTS2021 brain tumor dataset (b) Enhanced image by denoising (c) classification of feature parameters (d) bias-corrected image (e-f) enhanced segmentable images**

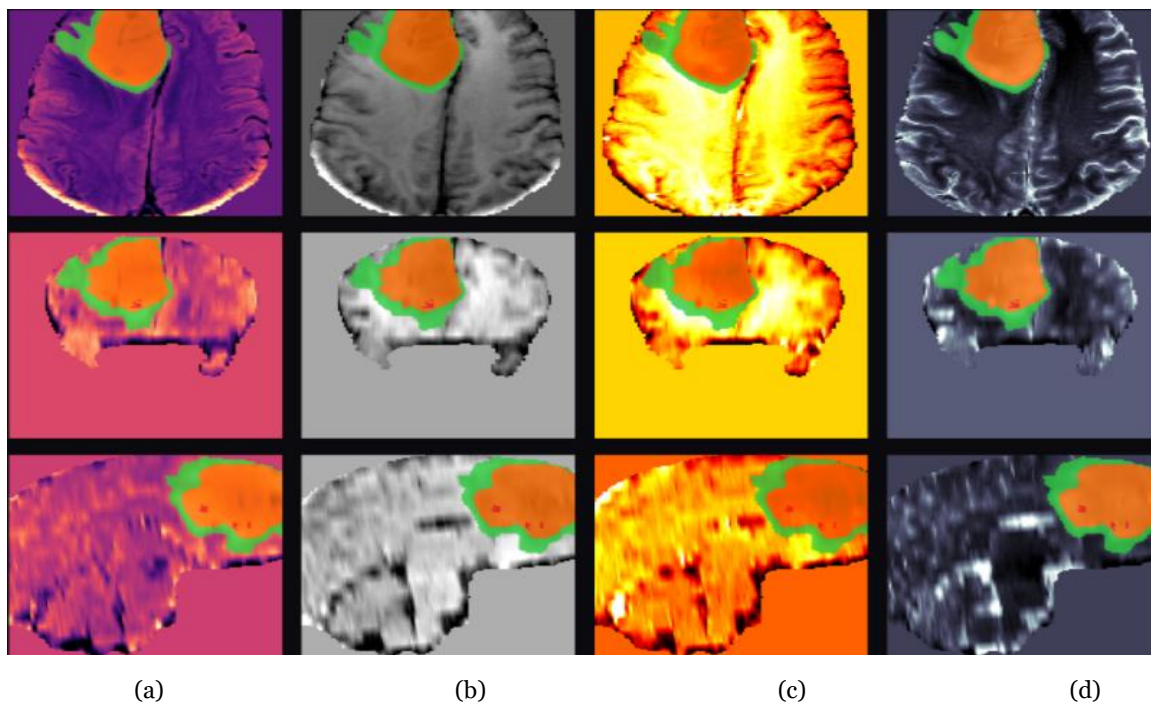
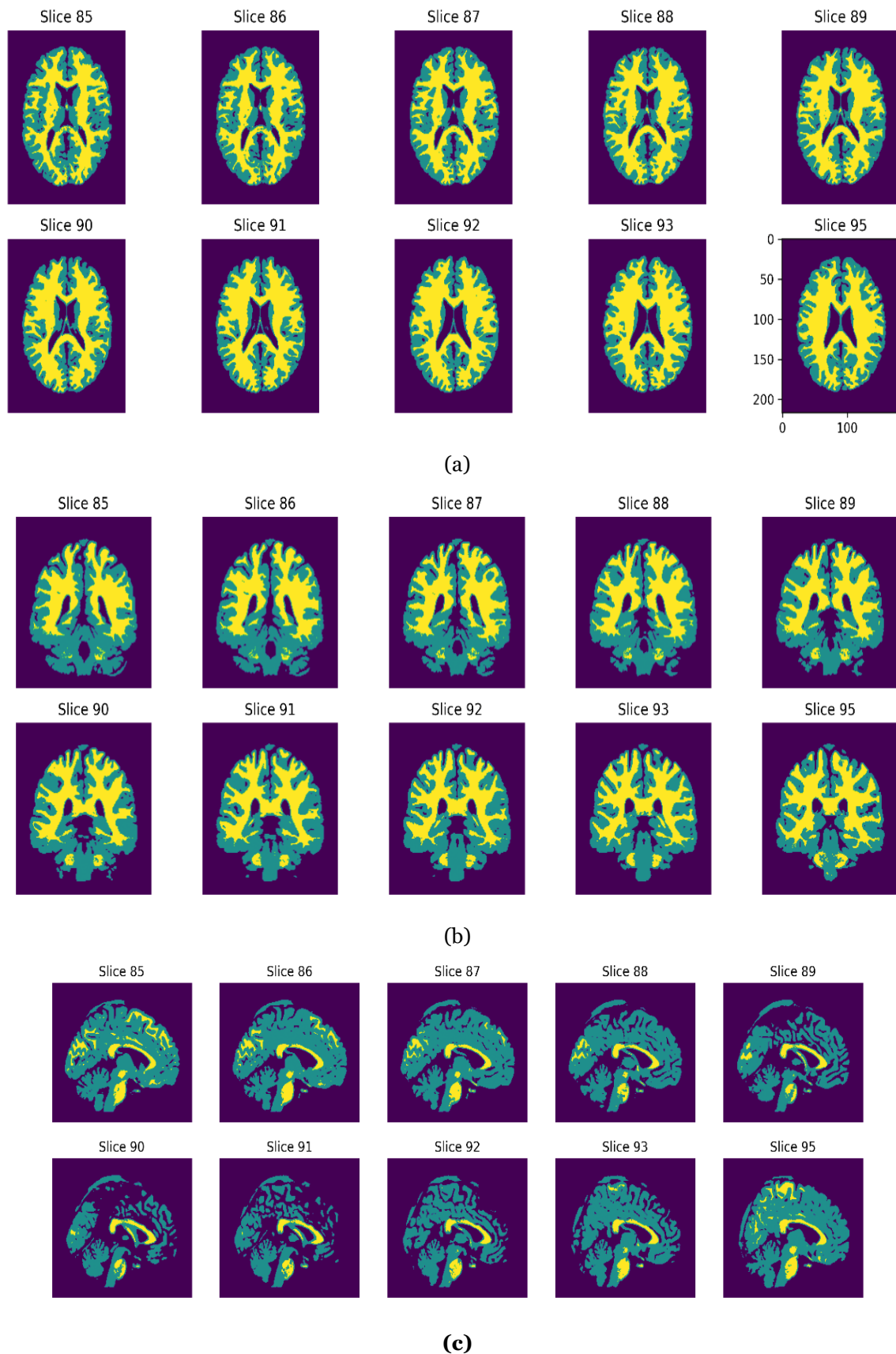
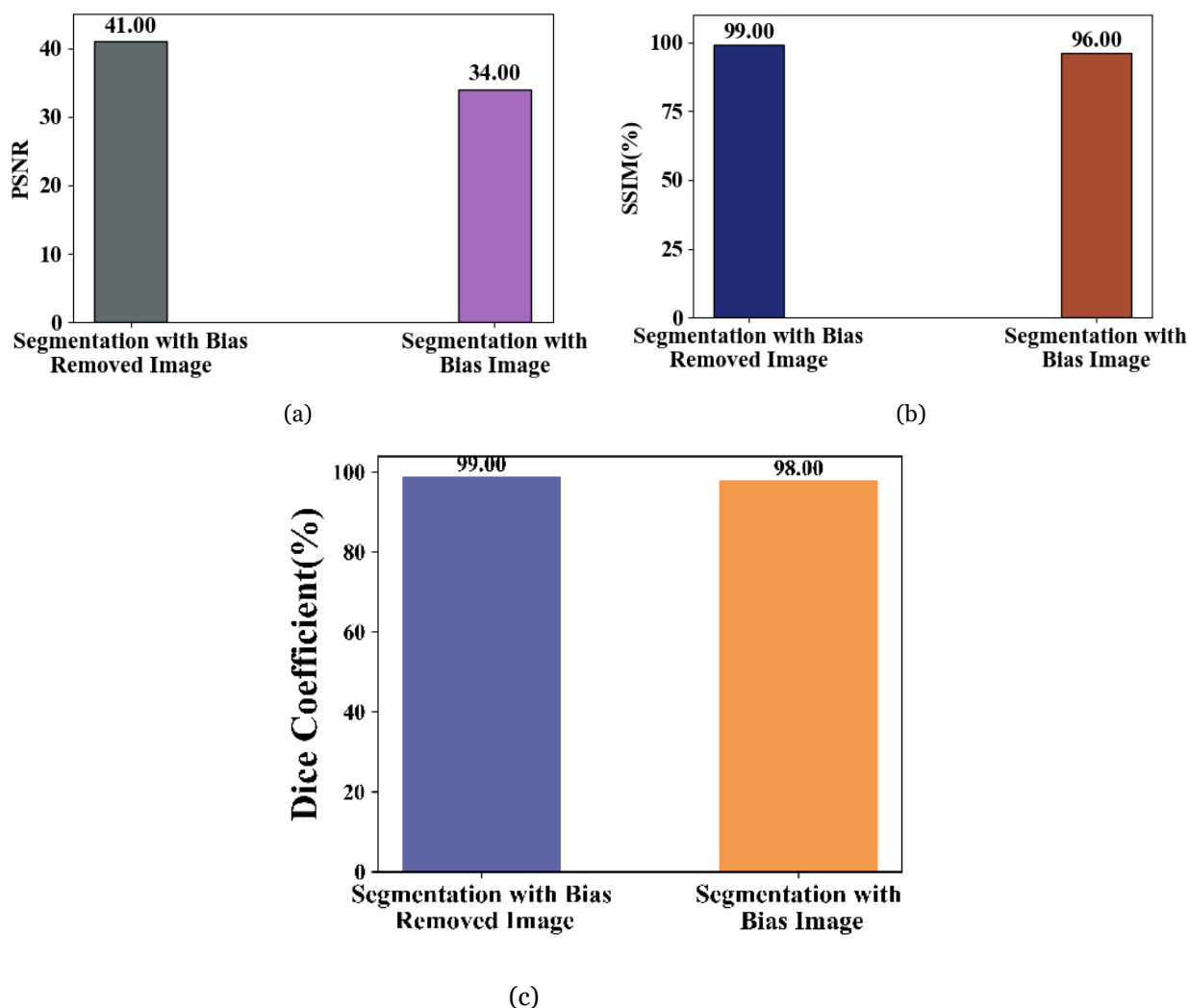


Fig. 8 Each rows contains axial, coronal and sagittal slices of BraTS2021 (a) tumor ground truth segmented images, (b) T1-w predicted segmented (c) T2-w predicted segmented; and (d) PD segmented slices.



**Fig. 9 Segmented slices of enhanced by bias-field correction and denoised MRI volume of a Brainweb T1w 3D (a) axial slices (b) coronal slices and (c) sagittal slices**



**Fig 10: Performance of BrainWeb dataset (a) PSNR (b) SSIM (c) Dice**

Figure 10 shows the performance evaluation on the BrainWeb dataset. The PSNR is 43.1771 for noise-removed images, higher than 35.8219 for noisy images. The SSIM is slightly better for noise-removed images at 99.8414% compared to 99.6553% for noisy images. The Dice score for bias-removed images is 99.7000%, outperforming 98.8700% for biased images, indicating better segmentation accuracy.

Figure 11 shows BraTS2012 dataset segmentation performance. This result indicates that segmentation with bias-removed images improves accuracy compared to biased images. The PSNR for bias-removed images is 42.24, significantly higher than the 31.46 in biased affected images. However, the SSIM is 99.25 for bias-removed images, higher than 93.50, showing better structural preservation. The Dice coefficient is 99.95 for bias-removed images versus 98.86 in bias field affected images which indicate the improved segmentation accuracy.

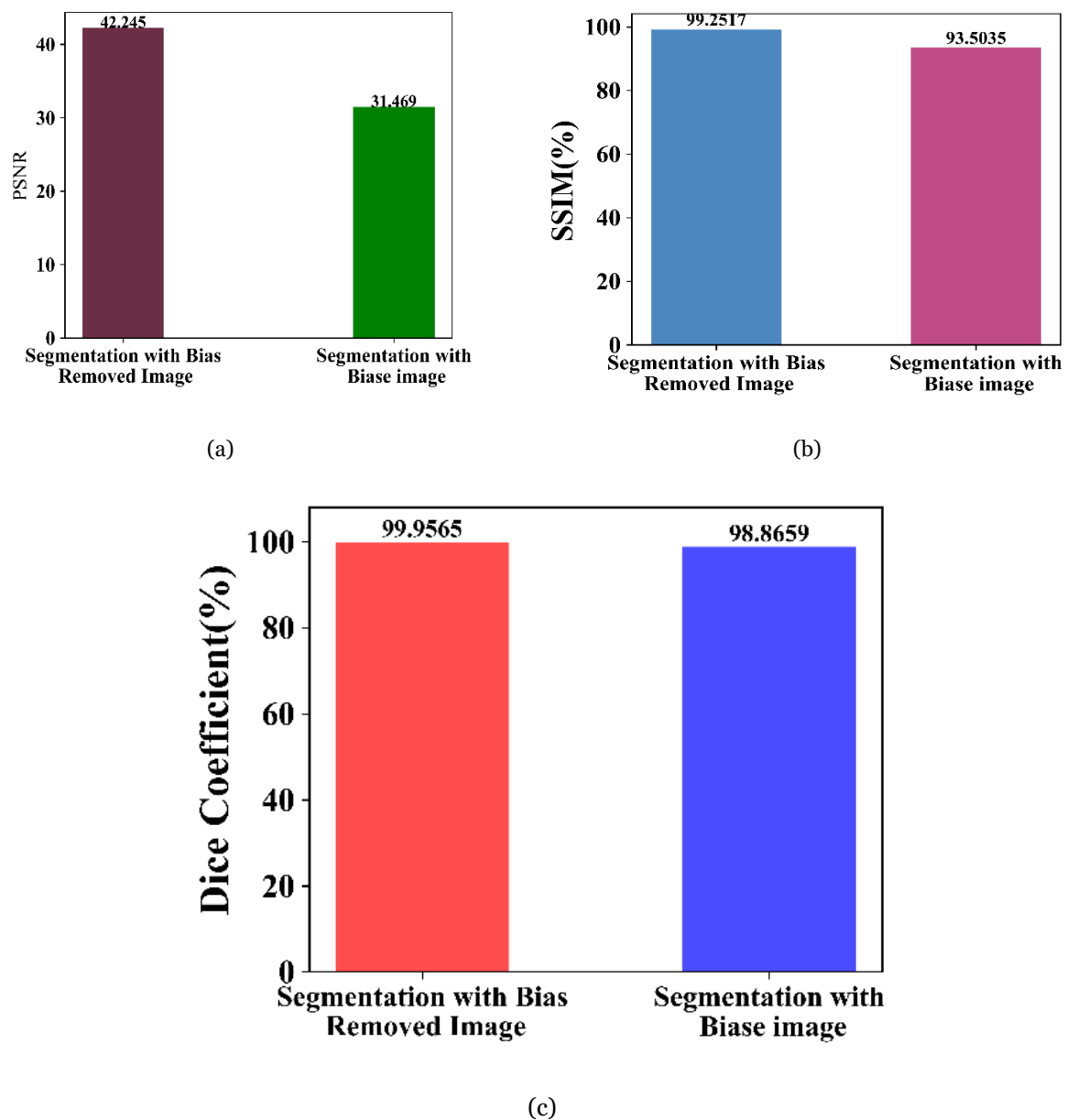


Fig 11: Performance of BT dataset (a) PSNR (b) SSIM (c) Dice

#### 4.4 Comparative analysis

Table 5: Average segmentation accuracy, measured by DC, achieved by existing methods and the proposed method using the BrainWeb dataset

| Noise N level (%) | IFLICM_IIN [36] | CEFAFCM [30] | esFCM [33] | Proposed |
|-------------------|-----------------|--------------|------------|----------|
| 0                 | 73.52           | 80.3         | 80.80      | 97.86    |
| 3                 | 85.49           | 85.65        | 85.65      | 99.2     |
| 5                 | 81.7            | 85.82        | 83.89      | 99.75    |
| 7                 | 82.63           | 81.75        | 82.16      | 98.98    |
| 9                 | 79.46           | 85.40        | 71.78      | 99.76    |

Table 6 presents DC scores, which indicate segmentation accuracy, for methods IFLICM\_IIN, CEFAFCM, esFCM and the proposed method using the BrainWeb dataset. Images with bias field effect at level 0% and noise level 0%, the segmentation accuracies are 73.5 (IFLICM\_IIN), 80.30 (CEFAFCM), 80.80 (esFCM) and 97.86 of proposed method that is significantly higher. Similarly with higher noise levels from 3-9%, the proposed methods should significantly higher performance than other existing methods as shown against each noise level for each method.

**Table 6: Average segmentation accuracy, measured by DC, achieved by existing methods and the proposed method using the BraTS2021 dataset**

| Noise level (%) | esFCM [33] | IFLICM_IIN [36] | EMT2FCM [38] | Proposed |
|-----------------|------------|-----------------|--------------|----------|
| 0               | 82.85      | 84.35           | 82.04        | 97.23    |
| 3               | 82.05      | 80.68           | 82.12        | 97.75    |
| 5               | 85.65      | 86.87           | 86.37        | 96.12    |
| 7               | 83.79      | 80.38           | 80.53        | 98.92    |
| 9               | 80.47      | 75.04           | 83.91        | 99.65    |

Table 7 displays the average segmentation accuracy measured by DC for existing methods (esFCM, IFLICM\_IIN, EMT2FCM) and the proposed method using the Brats2021 dataset. This assessment evaluates segmentation performance under different imaging conditions. The experimental denoising results obtained on noise levels (0%, 3%, 5%, 7%, 9%) where the proposed method consistently achieves superior accuracy of 98.54% compared to esFCM, IFLICM\_IIN and EMT2FCM.

**Table 7: Time complexity of proposed and existing methods on BrainWeb, IBSR and BT datasets**

| Datasets  | Methods         | Iterations |         |         |         |         |
|-----------|-----------------|------------|---------|---------|---------|---------|
|           |                 | 1          | 25      | 50      | 75      | 100     |
| BrainWeb  | IFLICM_IIN [36] | 5.16s      | 7.29 s  | 8.42 s  | 11.55 s | 13.74 s |
|           | CEFAFCM [30]    | 4.98s      | 8.79    | 10.69 s | 11.53 s | 14.40 s |
|           | esFCM [33]      | 6.16 s     | 25.75   | 28.65 s | 29.55 s | 44.98 s |
|           | Proposed        | 2.70 s     | 4.40 s  | 6.16 s  | 8.93 s  | 11.7 s  |
| BraTS2021 | esFCM [33]      | 16.24 s    | 24.65 s | 46.64 s | 42.93 s | 82.78 s |
|           | IFLICM_IIN [36] | 12.69 s    | 16.17 s | 25.68 s | 28.22 s | 30.51 s |
|           | EMT2FCM [38]    | 5.62 s     | 10.09 s | 14.76 s | 27.45 s | 30.75 s |
|           | Proposed        | 2.98 s     | 5.79 s  | 7.69 s  | 10.53 s | 14.40 s |
| IXI       | ORF2R2N [32]    | 3.08 s     | 6.10 s  | 9.11 s  | 6.16 s  | 18.93 s |
|           | ASTF [35]       | 13.53 s    | 14.40 s | 15.96 s | 23.45 s | 30.75 s |
|           | Proposed        | 2.7 s      | 4.61 s  | 6.45 s  | 9.54 s  | 10.52 s |

Table 8 presents the computational complexity of existing methods (IFLICM\_IIN, CEFAFCM, esFCM) and the proposed method across three different datasets. For example, on the BrainWeb dataset, at 50 iterations, the times are 8.42 s (IFLICM\_IIN), 10.69 s (CEFAFCM), 28.65 s (esFCM) and 6.16 s (proposed). Similarly, on the BraTS2021 dataset, at 75 iterations, the times are 25.93 s (IFLICM\_IIN), 46.24 s (esFCM), 15.96 s (EMT2FCM) and 10.53s. These results shown that the proposed method consistently demonstrates lower computational times compared to existing methods across datasets and iterations, highlighting its superiority in computational efficiency for image

segmentation tasks. Specifically, it achieves significantly faster execution times, such as 2.70 s on BrainWeb and 2.98s on BraTS2021, showcasing its advantage in reducing computational complexity. Comparative assessment with existing approaches involves various performance metrics include SSIM, Dice coefficient (DC), Accuracy, Precision, Sensitivity, Jaccard, Kappa score and Computational complexity. The comparative assessment is performed with previous techniques like CEFAFCM [30], esFCM [33-34], IFLICM\_IIN [36].

**Table 8: Performance comparison with exiting approaches on BrainWeb, BraTS2021 and IXI datasets**

| Datasets  | Techniques         | Accuracy (%) | Precision (%) | Sensitivity (%) | Jaccard (%) |
|-----------|--------------------|--------------|---------------|-----------------|-------------|
| BrainWeb  | DCNN [29]          | 85.53        | 86.70         | 85.70           | 92.50       |
|           | CEFAFCM [30-31]    | 94.49        | 91.16         | 92.44           | 93.10       |
|           | MICO [37]          | 93.39        | 91.97         | 89.75           | 92.2        |
|           | Proposed           | 99.26        | 98.62         | 99.53           | 95.87       |
| BraTS2021 | IFLICM_IIN [36]    | 92.56        | 92.20         | 93.62           | 92.58       |
|           | EMT2FCM [38]       | 93.07        | 94.49         | 95.97           | 94.96       |
|           | KBCFCM-LZM [39-40] | 92.57        | 91.17         | 94.62           | 95.25       |
|           | Proposed           | 99.29        | 99.56         | 99.56           | 96.66       |
| IXI       | ORF2R2N [32]       | 90.28        | 92.56         | 93.07           | 95.45       |
|           | ASTF [35]          | 94.86        | 94.62         | 94.94           | 92.64       |
|           | Proposed           | 99.91        | 99.91         | 99.91           | 98.27       |

Table 9 compares the performance metrics of existing techniques and the proposed method, DGN- HNG<sup>3</sup>, across BrainWeb, IBSR and BT datasets. For instance, on the BrainWeb dataset, the proposed method achieves 99.26% accuracy, 98.62% precision, 99.53% sensitivity, 95.8% Jaccard, outperforming DCNN, CEFAFCM and MICO. Similarly, on the BraTS2021 dataset, proposed method achieved and 99.29 % accuracy, 99.56% precision, 99.56% sensitivity and 96.6% Jaccard, surpassing IFLICM\_IIN, EMT2FCM and KBCFCM-LZM. In the IXI dataset, proposed method achieved and 99.91% accuracy, 99.91% precision, 99.91% sensitivity and 98.2% Jaccard, demonstrating superior performance compared to ORF2R2N and ASTF. Overall, performance of this method consistently shows higher accuracy and better segmentation results across all datasets, highlighting its superiority over existing methods in medical image analysis.

**Table 9: PSNR and MSE comparison with exiting approaches on BrainWeb, IBSR and BT datasets**

| Metrics         | BrainWeb     |          | BraTS2021    |          | IXI          |          |
|-----------------|--------------|----------|--------------|----------|--------------|----------|
|                 | EMT2FCM [38] | Proposed | EMT2FCM [38] | Proposed | EMT2FCM [38] | Proposed |
| Kappa score (%) | 93.65        | 98.38    | 92.57        | 96.24    | 88.54        | 98.74    |
| PSNR            | 41.25        | 43.1771  | 31.98        | 38.01    | 34.12        | 42.245   |
| MSE             | 0.0021       | 0.0015   | 0.0147       | 0.0137   | 0.0118       | 0.00095  |

Table 10 compares the Kappa score, PSNR, and MSE of different techniques on three brain datasets, showing that the proposed method outperforms others. It achieves the highest Kappa scores (98.38%, 96.24%, 98.74%), indicating superior segmentation accuracy. Similarly, PSNR values (43.1771, 38.01, 42.245) are the highest, reflecting better denoising quality, while its MSE values (0.0015, 0.0137, 0.00095) are the lowest, demonstrating minimal reconstruction error. Other method EMT2FCM show competitive but lower performance, confirming the effectiveness of the proposed approach.

## 6. CONCLUSION

The proposed patch-based diffusion model restores adequate quality images by denoising to detect anomalies in the healthy brain anatomy. It also allows the embedding of contextual information of individual microstructures and the tissue appearance of estimated brain anatomical structures. The quantitative analysis shows comparable results with the normal diffusion-based reconstruction methods and other non-patched methods in comparison with various baseline methods with significant high-performance differences on BrainWeb MRI dataset. The experimental results show that the image information around noisy regions can be used to replace the potential anomalies covered by noisy counterparts with the estimation of healthy brain structures. Unlike existing methods, which struggle with intense levels of noise, the introduced approach remains robust whether these artifacts occur in isolation or together. Extensive testing on MRI images using the three datasets demonstrates the superior accuracy, robustness and efficiency of the method compared to state-of-the-art techniques. It achieves 99.26% accuracy, 98.62% precision, 99.53% sensitivity, 95.8% Jaccard on BrainWeb dataset, 99.29 % accuracy, 99.56% precision, 99.56% sensitivity and 96.6% Jaccard on BraTS2021 dataset, 99.91% accuracy, 99.91% precision, and 99.91%. These findings underscore the resilience and efficacy of the proposed method in improving the accuracy and efficiency of MRI brain image segmentation. The initial values of the hyperparameter showed improved performance across the datasets. For generalization across a variety of anomalies, an ensemble can be employed in the model with different patch sizes and noise levels for further improvement in future studies. Despite its better reconstruction results and improved performance, this method has the drawback of a large inference time. In addition to employing a parallel computational framework, the denoising process can be improved by focusing on an efficient spatial context to reduce the inference time.

## REFERENCES

- [1] M. A. Bruno, E. A. Walker, and H. H. Abujudeh, 'Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction', *RadioGraphics*, vol. 35, no. 6, pp. 1668–1676, Oct. 2015, doi: 10.1148/rg.2015150023.
- [2] C. I. Bercea, B. Wiestler, D. Rueckert, and J. A. Schnabel, 'Towards Universal Unsupervised Anomaly Detection in Medical Imaging', Jan. 2024.
- [3] B. H. Menze et al., 'The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)', *IEEE Trans Med Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694.
- [4] S.-L. Liew et al., 'A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms', *Sci Data*, vol. 9, no. 1, p. 320, Jun. 2022, doi: 10.1038/s41597-022-01401-7.
- [5] H. J. Kuijf et al., 'Standardized Assessment of Automatic Segmentation of White Matter Hyperintensities and Results of the WMH Segmentation Challenge', *IEEE Trans Med Imaging*, vol. 38, no. 11, pp. 2556–2568, Nov. 2019, doi: 10.1109/TMI.2019.2905770.
- [6] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, 'Bayesian Skip-Autoencoders for Unsupervised Hyperintense Anomaly Detection in High Resolution Brain MRI', in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE, Apr. 2020, pp. 1905–1909. doi: 10.1109/ISBI45749.2020.9098686.
- [7] D. Shen, G. Wu, and H.-I. Suk, 'Deep Learning in Medical Image Analysis', *Annu Rev Biomed Eng*, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.
- [8] A. and P. N. and O. A. Q. Kascenas, 'Denoising Autoencoders for Unsupervised Anomaly Detection in Brain MRI', in *Proceedings of Machine Learning Research*, E. Konukoglu, B. Menze, A. Venkataraman, C. Baumgartner, Q. Dou, and S. Albarqouni, Eds., PMLR, Jul. 2022, pp. 653–664.
- [9] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, 'Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images', 2019, pp. 161–169. doi: 10.1007/978-3-030-11723-8\_16.
- [10] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, 'Scale-Space Autoencoders for Unsupervised Anomaly Segmentation in Brain MRI', 2020, pp. 552–561. doi: 10.1007/978-3-030-59719-1\_54.
- [11] X. Chen, S. You, K. C. Tezcan, and E. Konukoglu, 'Unsupervised lesion detection via image restoration with a normative prior', *Med Image Anal*, vol. 64, p. 101713, Aug. 2020, doi: 10.1016/j.media.2020.101713.
- [12] K. Sato, K. Hama, T. Matsubara, and K. Uehara, 'Predictable Uncertainty-Aware Unsupervised Deep Anomaly Segmentation', in *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2019, pp. 1–7. doi: 10.1109/IJCNN.2019.8852144.

- [13] W. H. L. Pinaya et al., 'Unsupervised brain imaging 3D anomaly detection and segmentation with transformers', *Med Image Anal*, vol. 79, p. 102475, Jul. 2022, doi: 10.1016/j.media.2022.102475.
- [14] F. Meissen, G. Kaissis, and D. Rueckert, 'Challenging Current Semi-Supervised Anomaly Segmentation Methods for Brain MRI', 2022, pp. 63–74. doi: 10.1007/978-3-031-08999-2\_5.
- [15] B. Nguyen, A. Feldman, S. Bethapudi, A. Jennings, and C. G. Willcocks, 'Unsupervised Region-Based Anomaly Detection In Brain MRI With Adversarial Image Inpainting', in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, IEEE, Apr. 2021, pp. 1127–1131. doi: 10.1109/ISBI48211.2021.9434115.
- [16] O. Özdenizci and R. Legenstein, 'Restoring Vision in Adverse Weather Conditions With Patch-Based Denoising Diffusion Models', *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 8, pp. 10346–10357, Aug. 2023, doi: 10.1109/TPAMI.2023.3238179.
- [17] M. D. A. R. F. Y. R. T. L. V. G. Andreas Lugmayr, 'RePaint: Inpainting using Denoising Diffusion Probabilistic Models', *ArXiv*, Jan. 2022.
- [18] F. and B. D. and K. J. and O. R. and S. A. Behrendt, 'Patched Diffusion Models for Unsupervised Anomaly Detection in Brain MRI', vol. 227. PMLR, 2024.
- [19] J. and J. A. and A. P. Ho, *Advances in Neural Information Processing Systems: Denoising Diffusion Probabilistic Models*, vol. 33. Curran Associates, Inc., 2020.
- [20] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4\_28.
- [21] N. J. Tustison et al., 'N4ITK: Improved N3 Bias Correction', *IEEE Trans Med Imaging*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010, doi: 10.1109/TMI.2010.2046908.
- [22] C. Baur, S. Denner, B. Wiestler, N. Navab, and S. Albarqouni, 'Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study', *Med Image Anal*, vol. 69, p. 101952, Apr. 2021, doi: 10.1016/j.media.2020.101952.
- [23] F. and B. M. and B. D. and K. J. and O. R. and S. A. Behrendt, 'Capturing inter-slice dependencies of 3D brain MRI-scans for unsupervised anomaly detection', in *Medical Imaging with Deep Learning*, 2022.
- [24] F. Meissen, G. Kaissis, and D. Rueckert, 'Challenging Current Semi-Supervised Anomaly Segmentation Methods for Brain MRI', 2022, pp. 63–74. doi: 10.1007/978-3-031-08999-2\_5.
- [25] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, 'f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks', *Med Image Anal*, vol. 54, pp. 30–44, May 2019, doi: 10.1016/j.media.2019.01.010.
- [26] P. and N. A. Dhariwal, 'Diffusion Models Beat GANs on Image Synthesis', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 8780–8794.
- [27] S. Elfving, E. Uchibe, and K. Doya, 'Sigmoid-weighted linear units for neural network function approximation in reinforcement learning', *Neural Networks*, vol. 107, pp. 3–11, Nov. 2018, doi: 10.1016/j.neunet.2017.12.012.
- [28] S. Raschka, 'MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack', *J Open-Source Softw*, vol. 3, no. 24, p. 638, Apr. 2018, doi: 10.21105/joss.00638.
- [29] Xu Y, Wang Y, Hu S, Du Y. Deep convolutional neural networks for bias field correction of brain magnetic resonance images. *The Journal of Supercomputing*. 2022 Nov;78(16):17943-68.
- [30] Gudise S, Giri Babu K, Satya Savithri T. An advanced fuzzy C-Means algorithm for the tissue segmentation from brain magnetic resonance images in the presence of noise and intensity inhomogeneity. *The Imaging Science Journal*. 2024 May 18;72(4):520-39.
- [31] Halder A, Talukdar NA. Kernel induced semi-supervised spatial clustering: a novel brain MRI segmentation technique. *Multimedia Tools and Applications*. 2023 Oct 28:1-29.
- [32] Shyamala B, Brahmananda SD. Brain tumor Classification using Optimized and Relief-based Feature Reduction and Regression Neural Network. *Biomedical Signal Processing and Control*. 2023 Sep 1;86:105279.
- [33] Jafrasteh B, Lubián-Gutiérrez M, Lubián-López SP, Benavente-Fernández I. Enhanced Spatial Fuzzy C-Means Algorithm for Brain Tissue Segmentation in T1 Images. *Neuroinformatics*. 2024 Apr 24:1-4.
- [34] Shukla V, Khandekar P, Khaparde A. Estimation of Nonhomogeneous Noise in 2D Magnetic Resonance Imaging. *International Journal of Imaging Systems and Technology*. 2022 Jul;32(4):1357-72.

- [35] Kumar R, Bhandari AK. Low-grade magnetic resonance image enhancement using adaptive sigmoid transformation function. *Health and Technology*. 2024 Mar;14(2):351-74.
- [36] Singh C, Ranade SK, Kaur D, Bala A. A novel approach for brain MRI segmentation and image restoration under intensity inhomogeneity and noisy conditions. *Biomedical Signal Processing and Control*. 2024 Jan 1;87:105348.
- [37] Jin R, Tong D, Chen Z. Level-set-based multiplicative intrinsic component optimization for brain tissue segmentation in T1-W and T2-W modality MRI. *Expert Systems with Applications*. 2023 Aug 15;224:119967.
- [38] Bose A, Maulik U, Sarkar A. An entropy-based membership approach on type-II fuzzy set (EMT2FCM) for biomedical image segmentation. *Engineering Applications of Artificial Intelligence*. 2024 Jan 1;127:107267.
- [39] Singh C, Ranade SK, Kaur D, Bala A. A kernelized-bias-corrected fuzzy C-means approach with moment domain filtering for segmenting brain magnetic resonance images. *Soft Computing*. 2024 Jan 8:1-25.
- [40] Zhao J, Liu Z, Sun Q, Li Q, Jia X, Zhang R. Attention-based dynamic spatial-temporal graph convolutional networks for traffic speed forecasting. *Expert Systems with Applications*. 2022 Oct 15;204:117511.