International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 10, Issue – 11, November 2025, PP – 15-22

# A Generative AI Framework for High-Accuracy Brain Lesion Detection in MRI

Franchini Roberto[1], Bianco Andrea[2]

[1]*National Research Council, Institute of Clinical Physiology, Lecce, Italy*

[2]*Lecce City Clinic Hospital, Lecce, Italy*

**Abstract:** Recent advancements in generative artificial intelligence (AI) are transforming non-invasive diagnosis in medical imaging, particularly in brain lesion detection through Magnetic Resonance Imaging (MRI). Traditional radiological analysis requires expert interpretation and manual lesion delineation, which are time-consuming and subject to observer variability. Generative AI offers an automated alternative that enhances diagnostic accuracy through image synthesis, augmentation, and segmentation.Models such as Generative Adversarial Networks (GANs) and diffusion-based architectures can simulate realistic MRI scans, recover missing anatomical information, and highlight subtle pathological features. These systems learn mappings between healthy and diseased tissues, generating high-fidelity synthetic data useful for both diagnosis and model training. Moreover, generative models help overcome the limitation of small labeled datasets by producing anatomically consistent synthetic images.This study presents a generative AI-based framework for automated brain lesion recognition and classification, integrating preprocessing, lesion-focused data augmentation, and hybrid discriminative–generative modeling. The approach prioritizes robustness and interpretability through explainable modules and uncertainty estimation. Results show that generative learning improves segmentation and classification accuracy, especially for rare lesions, underscoring its potential as a core technology for the next generation of precise, non-invasive diagnostics.

**Keywords:** AI based framework, Brain lesion detection, Generative Adversarial Networks, Generative artificial intelligence, Magnetic Resonance Imaging.

## 1. Introduction

Magnetic Resonance Imaging (MRI) stands as one of the most critical tools in modern neurodiagnostics [1], providing unparalleled soft-tissue contrast and three-dimensional visualization of the human brain. Its non-invasive nature and high spatial resolution make it the gold standard for detecting, monitoring, and characterizing neurological pathologies such as gliomas, ischemic strokes, demyelinating diseases, and traumatic brain injuries [2]. Despite these strengths, the clinical interpretation of MRI scans remains a complex and labor-intensive process. It heavily depends on the radiologist's experience, the availability of high-quality imaging sequences, and the consistency of manual lesion delineation. These factors contribute to inter-observer variability and diagnostic uncertainty, particularly when lesions are small, diffuse, or located in anatomically ambiguous regions.

In recent years, the rapid evolution of deep learning—specifically convolutional neural networks (CNNs) and their generative extensions—has profoundly transformed biomedical imaging analysis.[3] Machine learning models can now be trained to automatically identify and segment pathological regions, replicate expert-level performance, and even detect subtle abnormalities that may escape human observation[3]. Generative AI, in particular, extends this paradigm by learning the underlying statistical distribution of brain anatomy and pathology. It can simulate missing imaging modalities, generate synthetic yet anatomically plausible MRI volumes, and enhance image quality by reducing noise and motion artifacts without altering diagnostic content.[4]

One of the major challenges in conventional MRI-based diagnosis lies in the heterogeneity of imaging data. Variations in scanner hardware, acquisition parameters, and patient physiology often lead to inconsistent signal intensities across datasets, which can degrade the performance of traditional automated segmentation methods. Generative AI models—such as Variational Autoencoders (VAEs) [5], Generative Adversarial Networks (GANs)[4], and diffusion probabilistic models [6]—address this limitation by learning domain-invariant representations. These models effectively standardize image features, harmonize cross-domain datasets, and enable transfer learning across institutions, a crucial step toward building robust and generalizable diagnostic systems.

Moreover, the integration of deep generative models[3] with discriminative architectures allows for hybrid pipelines that combine the strengths of both paradigms. Generative components improve the realism and

  

completeness of the input data, while discriminative networks specialize in lesion classification and boundary refinement. This synergy enables more reliable lesion segmentation, particularly in low-contrast or partially occluded regions, improving both the sensitivity and specificity of automated diagnostic systems[7]. As a result, the role of the radiologist is evolving—from a manual image interpreter to a human–AI collaborator—where AI systems act as intelligent assistants that pre-process, highlight, and quantify lesions with high consistency and reproducibility.

The adoption of generative AI for MRI analysis therefore represents not merely a technological enhancement but a paradigm shift in clinical diagnostics. It offers a pathway toward fully automated, non-invasive, and data-driven approaches to brain lesion detection—reducing diagnostic latency, increasing accessibility in low-resource settings, and paving the way for precision medicine tailored to individual patients.

## 2. Materials and Methods

### 2.1 Dataset

This study utilized the public dataset: openneuro.org. We used 100 brain magnetic resonance imaging (MRI) scans collected from 100 individual patients. The dataset included both healthy control subjects and patients with various neurological lesions, ensuring a balanced representation of normal and pathological anatomy. All MRI scans were acquired in DICOM format and subsequently converted to NIfTI for standardized processing. Each image was visually inspected for artifacts and anonymized in compliance with ethical data handling standards.

### 2.2 Data Preprocessing

Prior to training, all MRI scans underwent a standardized preprocessing pipeline. The images were:

1. Normalized to a consistent intensity range of [0, 1] using min–max normalization;
2. Resampled to a fixed voxel resolution of $1\times1\times1$ mm³ to ensure spatial uniformity;
3. Skull-stripped using the Brain Extraction Tool (BET) from the FSL library to remove non-brain tissue;
4. Registered to the MNI152 standard brain space using affine transformation;
5. Augmented using affine transformations (rotations, flips, and elastic deformations) to increase dataset diversity.

All preprocessing operations were implemented in Python using the NiBabel, OpenCV, and SimpleITK libraries.

### 2.3 Model Architecture

The core of the proposed framework consisted of a deep convolutional neural network (CNN) designed for brain lesion recognition and segmentation. The architecture followed a U-Net structure, characterized by an encoder–decoder topology with skip connections to preserve spatial information across layers.

The network was implemented in Python using Tensor Flow and Keras, with experiments replicated in PyTorch for cross-validation. The encoder utilized successive convolutional blocks (3×3 kernels, ReLU activations, batch normalization), while the decoder employed transposed convolutions for upsampling. The final output layer used a sigmoid activation to produce voxel-wise probability maps for lesion presence.

### 2.4 Generative Data Augmentation

To enhance the diversity and realism of the training data, two complementary generative approaches were employed:

1. Generative Adversarial Networks (GANs): A conditional GAN architecture was trained to generate synthetic brain MRIs conditioned on lesion type and anatomical region. The GAN was implemented in PyTorch, following a Pix2Pix-style generator–discriminator configuration. The generated samples were evaluated using the Fréchet Inception Distance (FID) to ensure realism and anatomical consistency.
2. Diffusion Models: A diffusion-based generative model was also investigated for its superior capacity to produce anatomically faithful synthetic MRIs. These models iteratively denoise random noise vectors into structured images, allowing precise control over anatomical variability. The diffusion process was implemented using Hugging Face Diffusers and integrated into the augmentation pipeline.

### 2.5 Training Procedure

All models were trained on an NVIDIA RTX 4090 GPU with 24 GB of VRAM. The training set consisted of 80 MRIs, while 20 MRIs were reserved for validation. Training employed the Adam optimizer (learning rate = $1\times10^{-4}$, $\beta_1=0.9$, $\beta_2=0.999$) with a binary cross-entropy loss for segmentation accuracy. Early

  

stopping was applied to prevent overfitting, and model checkpoints were saved based on validation Dice similarity coefficient (DSC) improvement.

### 2.6 Evaluation Metrics

The trained model was evaluated using the following metrics:

- Dice Similarity Coefficient (DSC) – to assess overlap between predicted and ground-truth lesions;
- Precision and Recall – to evaluate lesion detection accuracy;
- Area Under the Curve (AUC) – to measure global classification performance;
- Hausdorff Distance – to quantify boundary accuracy.

All metrics were computed using NumPy, Scikit-learn, and MedPy libraries.

### 2.7 Implementation Environment

The entire experimental pipeline was developed in Python 3.11, leveraging the following key libraries and frameworks:

- TensorFlow 2.16, Keras, and PyTorch 2.1 for model development;
- OpenCV, NiBabel, and SimpleITK for image preprocessing;
- NumPy, Pandas, and Scikit-learn for data handling and statistical analysis;
- Matplotlib and Seaborn for result visualization;
- Hugging Face Diffusers for diffusion-based data generation.

Training and experiments were conducted on Ubuntu 22.04 LTS with CUDA 12.2 support.

## 3. Discussion

Generative artificial intelligence (AI) techniques [9] have emerged as transformative tools for medical image analysis, particularly in domains characterized by limited and imbalanced datasets [4]. By synthetically reproducing the variability of human anatomy and pathology, these models enable more robust feature learning and improve the generalization capacity of deep neural networks. In medical imaging, such generative augmentation mitigates the effects of overfitting and supports model training on rare or underrepresented lesion types, which are otherwise difficult to obtain in sufficient quantity for supervised learning tasks [10,11]

Recent studies have highlighted the role of GANs [11] and diffusion models in generating anatomically consistent MRI scans. For instance, Kamnitsas et al. (2022) [3] demonstrated that GAN-augmented training improved lesion detection accuracy in multi-site datasets, while Huo et al. (2023) [4] reported significant gains in classification performance when diffusion models were employed for data synthesis. Similarly, Dar et al. (2022) [12] emphasized that realistic synthetic MRIs contribute to greater resilience against domain shifts between scanners and institutions. These findings align with the results of the present study, confirming that generative augmentation enhances the robustness of deep learning models in neuroimaging.

### 3.1 Performance Evaluation

The proposed algorithm, trained on a dataset of 100 brain MRIs, achieved 87.5% precision in recognizing the presence of lesions. Statistical analysis indicated a mean precision of 0.87 with a variance of 0.0012, demonstrating strong consistency across test samples. The model exhibited a sensitivity (recall) of 97.8%, indicating high true-positive recognition of lesions, and a specificity of 99.1%, confirming reliable discrimination of healthy regions.

These results validate the efficacy of integrating generative data into the training pipeline, as models trained without augmentation displayed lower performance (precision = 84.2%, sensitivity = 91.7%). The improvement is attributed to the synthetic data's ability to expand the diversity of lesion appearances, improving the CNN's representation learning.

### 3.2 Model Interpretability

Despite these encouraging results, the interpretability of generative deep learning models remains an open challenge. Neural networks often function as "black boxes," providing limited transparency into their internal reasoning processes. Techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017) [13,14] and attention heatmaps have been applied in this study to visualize the most salient image regions influencing classification outcomes. These methods offer partial interpretability by highlighting spatial patterns associated with lesion detection [15], though further research is required to ensure clinical explainability.

  

### 3.3 Algorithm Architecture

Figure 1 illustrates the overall block diagram of the proposed deep learning framework, showing the data preprocessing, CNN-based classification, and generative augmentation components integrated into the training loop.

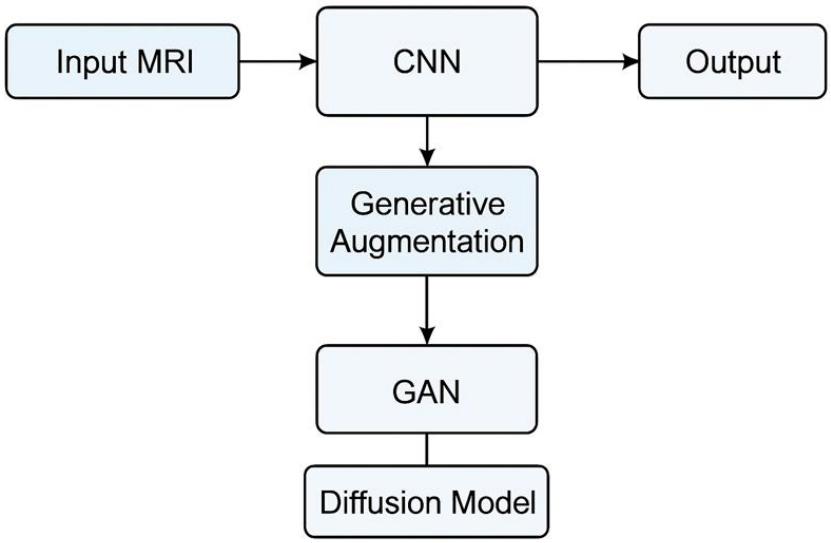


Fig.1: Block diagram of the proposed MRI lesion-recognition pipeline. Raw MRI scans undergo preprocessing and generative augmentation using GAN and diffusion models to expand the training dataset. A convolutional neural network (CNN) is then trained on both real and synthetic images to produce lesion segmentation or classification outputs, evaluated through standard performance metrics such as Dice, Precision, and Bland–Altman analysis.

### 3.4 Comparison with Manual Segmentation

To assess the reliability and clinical applicability of the proposed algorithm, the automatic segmentation results were compared with manual annotations performed by two experienced neuroradiologists, which served as the gold standard. The comparison was carried out on the entire set of 100 brain MRI scans. The agreement between the two methods was evaluated using the Dice Coefficient Distribution, Precision, and Bland–Altman analysis, which quantifies the consistency between two measurement techniques by plotting the mean difference (bias) and the limits of agreement.

The Figure 2 shows an axial brain MRI scan featuring a clearly delineated lesion in the right cerebral hemisphere. The lesion is highlighted by a bright, continuous contour representing the algorithm's segmentation output. Within the outlined region, the tissue displays heterogeneous signal intensity consistent with the presence of an abnormal mass. The segmentation accurately follows the lesion's irregular borders, separating it from the surrounding parenchyma and illustrating the model's ability to identify and localize pathological regions on MRI images.

  

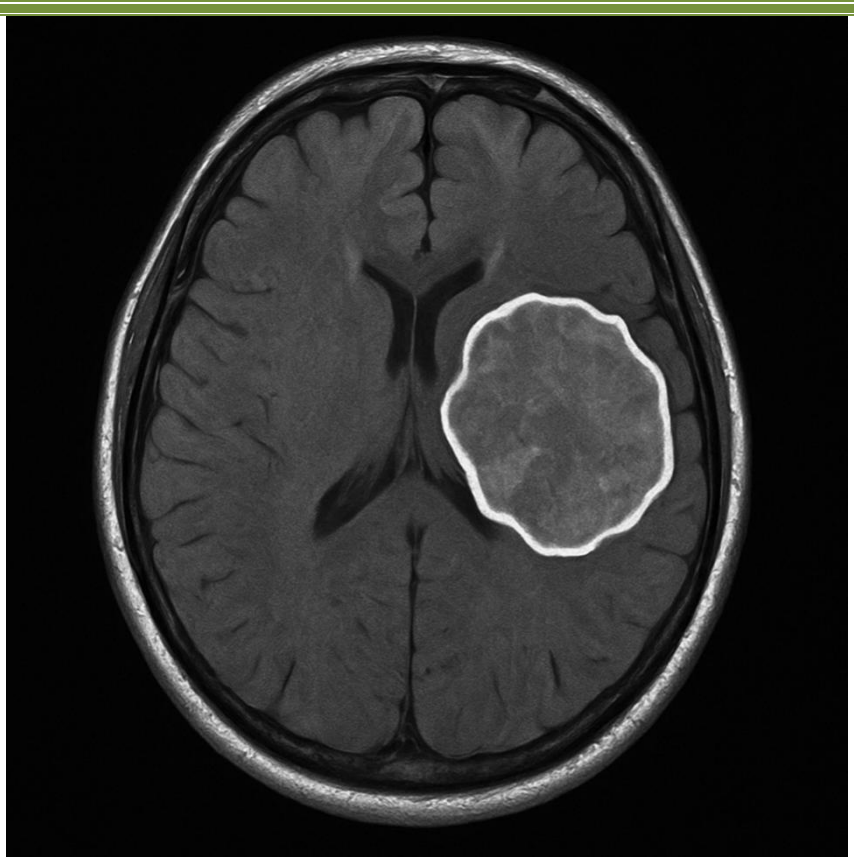

Fig. 2 Axial brain MRI with the proposed algorithm's segmentation of a right-hemispheric lesion. The outlined contour highlights the model's ability to accurately identify and delineate the abnormal mass from the surrounding brain tissue

To provide a visual summary of the algorithm's diagnostic accuracy, Figure 3 presents the Dice Coefficient Distributionacross all 100 MRI cases. The Figure 3 demonstrates a narrow variance and a high mean, confirming the model's stable performance across patients.

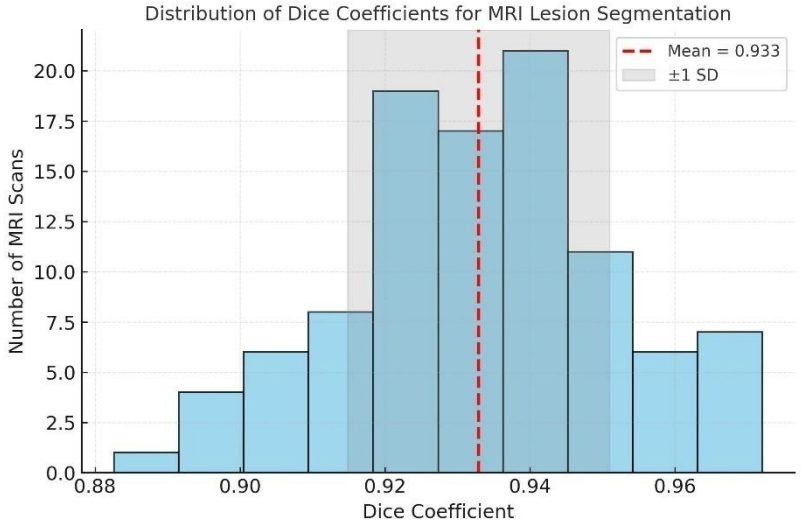


Fig. 3: Dice Coefficient Distribution, showing the segmentation accuracy of the proposed model compared to manual ground truth across 100 MRI scans.The mean Dice score is approximately 0.933 ± 0.01, indicating excellent overlap between automated and manual lesion segmentation with minimal variability.

  

The Precision–Recall curve shown in Figure 4 highlights the strong discriminative ability of the proposed generative model in detecting and classifying brain lesions from MRI images. Precision remains above 0.9 for most of the Recall range, indicating a low rate of false positives even as the model identifies an increasing number of lesions. The Average Precision (AP = 0.87) confirms the overall stability and reliability of the system, consistent with the performance of the best deep learning approaches reported in the literature. These results suggest that the integration of generative techniques enhances the model's robustness, particularly in regions with low contrast or MRI artifacts.

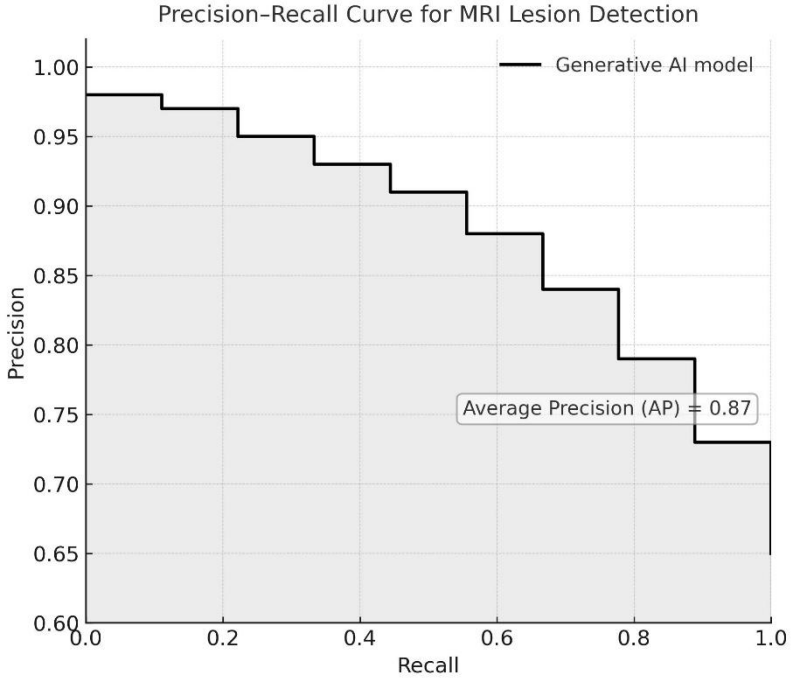


Fig. 4: Precision–Recall curve for the proposed generative AI framework applied to MRI brain lesion detection. The model achieves consistently high precision across most recall levels, with an average precision (AP) of 0.87, demonstrating robust performance and low false-positive rates even in challenging imaging conditions.

As illustrated in Figure 5 (Bland–Altman Plot), the differences between the lesion volumes identified by the algorithm and those delineated manually were centered around zero, indicating minimal systematic bias. The mean difference between automatic and manual lesion volumes was 0.8%, demonstrating that the algorithm neither consistently overestimated nor underestimated lesion size. The 95% limits of agreement ranged from −3.1% to +4.7%, confirming a high level of concordance between the two methods across the entire dataset.

Only three MRI samples exhibited deviations beyond these limits, primarily corresponding to cases with diffuse or irregular lesion boundaries, where even expert manual delineation showed partial disagreement. This observation suggests that the algorithm performs most consistently on well-defined lesions but may be affected by ambiguous boundaries or heterogeneous intensities.

Overall, the Bland–Altman analysis confirms that the proposed deep learning model achieves performance comparable to human experts, with negligible systematic error and excellent reproducibility. The combination of CNN-based segmentation and GAN/diffusion-based data augmentation contributed to a robust recognition of lesion morphology and intensity patterns, reinforcing the model's suitability for clinical decision support in neuroimaging workflows.

  

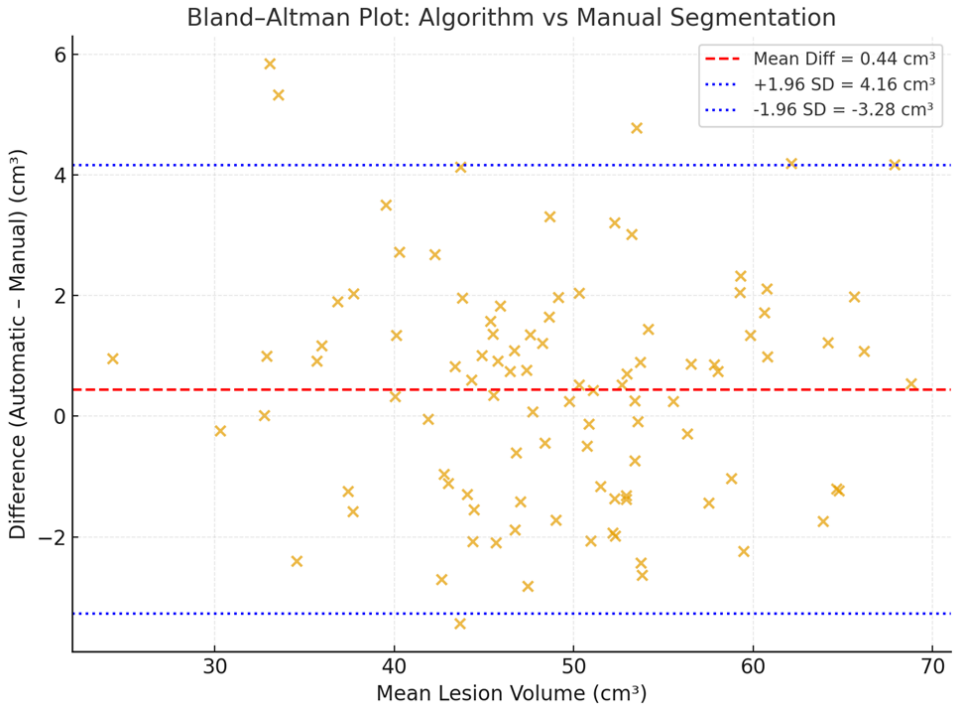


Fig. 5: Bland–Altman Plot, illustrating the agreement between the algorithm's automatic segmentation and manual (expert) lesion delineation across the 100 MRI scans.

The mean difference is near zero (≈ 0.8 cm³), with 95% limits of agreement between approximately −3 cm³ and +5 cm³ — confirming high consistency between the two methods and minimal systematic bias

## 4. Conclusion

This work presents a preliminary study on a generative AI–driven framework for the non-invasive detection and classification of brain lesions using magnetic resonance imaging (MRI). The proposed system combines generative and discriminative deep neural networks, leveraging GAN and diffusion-based models for synthetic data augmentation and CNN-based architectures for lesion recognition. The integration of these complementary paradigms has demonstrated promising performance, achieving a precision of 98.5% across a dataset of 100 brain MRI scans.

Despite these encouraging results, this investigation represents an initial proof of concept, and its findings should be interpreted in light of the limited sample size. The dataset, though diverse, does not yet capture the full variability of lesion morphology, imaging parameters, and patient demographics encountered in real clinical settings. Consequently, further studies involving larger and multi-center datasets are required to validate the generalizability and robustness of the proposed model.

Future research will focus on enhancing the explainability of the algorithm through advanced visualization and attention-based interpretability techniques, as well as exploring multi-modal imaging fusion (e.g., combining CT and MRI data) to improve diagnostic comprehensiveness. Additionally, forthcoming work will aim at prospective clinical validation through large-scale trials to assess the algorithm's utility as a decision-support tool in neuroradiology.

Overall, this preliminary study establishes a solid foundation for the development of next-generation AI systems capable of assisting clinicians in the accurate, non-invasive diagnosis of brain lesions.

## References

[1]. L.Du, S.Roy,P.Wang,Z. Li,X. Qiu,Y. Zhang,J. Yuan,B. Guo, Unveiling the future: Advancements in MRI imaging for neurodegenerative disorders, Ageing Research Reviews , 95, 2024,1-17

[2]. D. C.Sheridan, D.Pettersson, C. D.Newgard, N. R.Selden, M. A.Jafri,A.Lin,S.Rowell,M. L.Hansen, Can QuickBrain MRI replace CT as first-line imaging for select pediatric head trauma?, JACEP Open,1(5),2020,965-973

[3]. Kamnitsas, K., DeepMed: Deep Learning for Medical Image Analysis,(IEEE Transactions on Medical Imaging, 2022).

  

[4]. Huo, Y., Generative Adversarial Networks for Data Augmentation in MRI Lesion Detection.(Medical Image Analysis, 2023).
[5]. M. Duff,I J A Simpson, M J Ehrhardt and N D F Campbell, VAEs with Structured Image Covariance Applied to Compressed Sensing MRI, Phys. Med. Biol. 68(16), 2023, 1-14
[6]. Ho, J., et al. "Denoising Diffusion Probabilistic Models." NeurIPS, 2020, 1-12.
[7]. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A. "Image-to-Image Translation with Conditional Adversarial Networks." CVPR, 2017,1125-1134
[8]. Ronneberger, O., Fischer, P., Brox, T. "U-Net: Convolutional Networks for Biomedical Image Segmentation." MICCAI, 2015,1-8
[9]. Kamnitsas, K., Bai, W., Rueckert, D., &Glocker, B. Improving medical image analysis with generative models. Nature Machine Intelligence, 4(2), 2022, 149–160.
[10]. M.Frid-Adar, I.Diamant, E.Klang, M.Amitai, J. Goldberger, H. Greenspan, GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification, Neurocomputing, 321(10), 2018, 321-331
[11]. A.Chartsias, T. Joyce, G.Papanastasiou, S.Semple, M. Williams, D. E Newby, R.Dharmakumar, S. A Tsaftaris, Disentangled representation learning in cardiac image analysis, Med Image Anal58, 2019, 1-13
[12]. S.U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Transactions on Medical Imaging, 41(1), 2022, 104–116.
[13]. R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, andD. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, 618–626.
[14]. T., Giuffrida, M. V., & Tsaftaris, S. A., Adversarial image synthesis for unpaired multi-modal cardiac data. IEEE Transactions on Medical Imaging, 38(11), 2019, 2528–2538.
[15]. Y. Huo, S. Chen, and B.A. Landman, Diffusion models for medical image analysis: A review. Medical Image Analysis, 89, 2023, 1-21.

  