Cluster Computing (2026) 29:148
https://doi.org/10.1007/s10586-025-05893-x

Evaluating augmented deep networks for MRI-based brain anomaly classification

M. Mücahit Enes Yurtsever[1] · Süleyman Eken[1] · Bilgehan Arslan[2] · Yılmaz Atay[2] · Şeref Sağıroğlu[2] · Emrah Celtikci[3] · Alp Ozgun Borcek[3] · Alperen Sozer[4] · M. Caglar Sahin[5] · Zuhal Demirtas[6] · Kerem Nernekli[7]

Received: 14 April 2025 / Revised: 29 November 2025 / Accepted: 22 December 2025
© The Author(s) 2026

Abstract
Accurate and early detection of brain anomalies using MRI is critical for effective diagnosis and treatment planning. Recent advances in deep learning, particularly convolutional neural networks, have significantly improved the capabilities of automated medical image analysis, enabling more precise and scalable solutions in clinical settings. However, the performance of such models is highly dependent on the quality, diversity, and annotation detail of the training datasets. Building on this foundation, in this study, we investigate the performance of augmented deep neural networks for MRI-based brain anomaly classification using the BraTS 2021 dataset and our newly introduced Gazi Brains 2025 dataset, which includes MRI scans from 500 patients, a substantial portion of which are annotated for various brain anomalies. The dataset supports both binary (tumor versus normal) and multiclass (seven neurological conditions) classification tasks. We develop and evaluate a series of deep learning and transformer-based models for anomaly classification, with a particular focus on the impact of synthetic data augmentation. Using StyleGANv3 and Guided Diffusion, we generate synthetic MRI scans to enhance training data diversity and examine their effect on model performance. In addition, a ß-VAE–based generative pipeline is employed to further expand the synthetic dataset, providing controllable latent representations and additional variability through VAE sampling. Experimental results show that all three augmentation strategies-StyleGANv3, Guided Diffusion, and ß-VAE- significantly improve classification accuracy compared to baseline models trained solely on the original dataset. DenseNet achieved an accuracy value of up to 91% in binary classification when trained with augmented data, and EfficientNetV2S also achieved up to 72% in multiclass classification. Although StyleGANv3-generated images exhibited superior visual quality (low FID scores), the method was limited in sample volume. In contrast, the diffusion-based approach allowed the creation of larger synthetic datasets, though at the cost of extended training and sampling times. The ß-VAE model produced a moderate number of anatomically coherent samples with lower computational cost, offering a balanced alternative between quality and scalability.

Keywords Brain MRI classification · Deep learning · Synthetic data augmentation · Neuro-anomaly detection · Medical image analysis

1 Introduction

Brain tumors, which include a diverse range of neoplasms within the central nervous system, pose significant health challenges worldwide. Based on a projected annual growth rate of 2.5%, the estimated global incidence of malignant brain and central nervous system (CNS) tumors in 2025 is approximately 403,564 new cases, with 285,578 deaths. In particular, these figures exclude non-cancerous brain tumors, which represent just over half (53%) of the incidence in the UK. For the United States, the projected numbers for 2025 are approximately 26,066 new cases and 19,951 deaths[1]. These statistics underscore the substantial impact of brain tumors on public health systems and the urgent need for effective diagnostic and therapeutic strategies[2].

[1] https://www.thebraintumourcharity.org/

[2] https://www.cancer.org/cancer/types/brain-spinal-cord-tumors-adults/about/key-statistics.html

Extended author information available on the last page of the article

Published online: 12 February 2026

Early detection of brain tumors is vital to improve patient outcomes. Detection in time allows for rapid intervention, potentially improving survival rates and quality of life. For example, the relative survival rate for patients with malignant brain tumors is approximately 35.7%; however, this rate varies significantly depending on the type of tumor and the stage of diagnosis[3]. Early-stage detection often enables a broader range of treatment options, including surgical resection, radiation therapy, and chemotherapy, which can be more effective when the tumor is localized. Conversely, delayed diagnosis may result in limited treatment possibilities and poorer prognoses.

The global burden of brain and CNS cancers has exhibited an increasing trend over recent decades. Data from the Global Cancer Observatory indicate that the incidence of these malignancies has risen, with significant geographical and regional variations [1]. Specifically, a five-fold difference in incidence rates has been observed between regions with the highest rates, predominantly in Europe, and those with the lowest, mainly in Asia. This variability may reflect differences in diagnostic capabilities, reporting practices, environmental exposures, and genetic predispositions[4]. Addressing these disparities through improved diagnostic infrastructure and increased awareness is crucial in mitigating the global impact of brain tumors.

Brain tumor abnormalities encompass a diverse range of neoplastic growths that originate in the brain or spread from other parts of the body. These abnormalities can be broadly classified into benign and malignant tumors, with the latter posing a significant threat due to their aggressive nature and potential for rapid progression [2]. Common malignant brain tumors include glioblastomas (GBM), high-grade gliomas (HGG), low-grade gliomas (LGG), metastatic brain tumors (MET), and meningiomas (MNG), among others. Some tumors, such as schwannomas (SCH) and post-treatment effects (POST) mimicking tumors, can be challenging to differentiate from malignant cases, complicating diagnosis. Symptoms of brain tumors vary depending on their size, location, and rate of growth but often include persistent headaches, seizures, cognitive impairment, and motor dysfunction [3, 4]. Advanced imaging techniques, such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans, play a crucial role in detecting these abnormalities, while deep learning-based automated classification methods are increasingly being explored to enhance diagnostic accuracy and efficiency.

MRI scans are fundamental in detecting and diagnosing brain tumor abnormalities, providing critical insights into their location, size, and structural impact [5]. MRI scans, utilizing strong magnetic fields and radio waves, offer high-resolution, multi-plane images that reveal soft tissue contrast with exceptional clarity, making them the gold standard for brain tumor detection. Specialized MRI techniques such as functional MRI (fMRI), diffusion-weighted imaging (DWI), and perfusion MRI further enhance diagnostic precision by analyzing tumor metabolism, blood flow, and cellular density [6]. These imaging modalities allow clinicians to distinguish between different tumor types, assess malignancy levels, and monitor disease progression, thereby aiding in treatment planning and surgical decision-making.

The automated classification of brain MRI scans is a crucial task in medical imaging, aiding in the early detection and diagnosis of neurological disorders. Traditional diagnostic methods rely heavily on expert radiologists, making the process time-consuming and prone to subjective variability. With the rapid advancements in deep learning, automated approaches have demonstrated significant potential in improving diagnostic accuracy and efficiency. However, challenges persist in distinguishing pathological cases from normal scans, especially when dealing with diverse and complex anomalies [7, 8]. This study aims to address these challenges by leveraging modern deep learning architectures and synthetic data augmentation techniques to enhance classification performance.

One of the primary difficulties in automated MRI classification is the limited availability of high-quality, labeled datasets. To overcome this, researchers have explored synthetic data augmentation techniques, which enhance model robustness by artificially expanding the dataset [9, 10]. This study utilizes the Gazi Brains 2025 dataset, a recently developed collection of brain MRI scans, to evaluate the impact of synthetic data on model performance. By systematically applying data augmentation, the study aims to improve generalization and reduce the risk of overfitting, which is a common issue in deep learning-based medical imaging tasks.

Compounding these difficulties is the perennial challenge of data scarcity in the medical domain, where the acquisition of large, annotated datasets is often constrained by privacy concerns, costly expert labeling, and the relative rarity of certain pathological conditions. Synthetic data augmentation emerges as a powerful strategy to mitigate these limitations by artificially expanding the diversity and size of training datasets. By generating realistic yet novel medical images, these techniques can help in creating more robust and generalized deep learning models. This is particularly vital for improving model performance on underrepresented classes and enhancing its ability to discern subtle anatomical variations that signify disease, thereby ensuring that automated systems are trained on a comprehensive spectrum of clinical

[3] https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/

[4] https://www.iarc.who.int/reference/cancers-of-the-brain-and-cns-global-patterns-and-trends-in-incidence

presentations. The strategic use of synthetic data thus serves as a critical enabler, allowing for the development of more reliable and accurate diagnostic tools without solely relying on the procurement of extensive real-world patient data.

In addition to dataset augmentation, evaluating the generalizability of classification models across heterogeneous datasets is essential for validating their clinical applicability. In this study, particular emphasis is placed on the Gazi Brains 2025 and BraTS 2021 dataset[5], a recently compiled collection of brain MRI scans that reflects region-specific clinical imaging characteristics. To assess cross-dataset robustness, the model is also evaluated on the widely recognized BraTS 2021 dataset, a benchmark in brain tumor segmentation and classification. Both datasets undergo an identical augmentation pipeline, enabling a controlled comparative analysis. This dual-dataset evaluation framework not only highlights the model's adaptability but also provides critical insights into how augmentation strategies influence diagnostic performance, with a specific focus on the clinical relevance of the Gazi Brains 2025 dataset.

The classification task in this study focuses on distinguishing pathological cases from normal brain MRI scans. The pathological class comprises seven distinct anomalies: CVM, HGG, LGG, MET, MNG, POST, and SCH. The inclusion of multiple anomaly types introduces additional complexity, as the classification model must generalize across a diverse set of pathological conditions. Deep learning architectures, including convolutional neural networks (CNNs), vision transformers, and diffusion-based models, are explored to determine their efficacy in handling these variations.

A principal contribution of this study is the in-depth examination of the impact of data augmentation techniques on the robustness, reliability, and generalization capacity of deep learning-based brain MRI classification models. Through a systematic evaluation of diverse augmentation strategies, the research aims to delineate the most effective methods for optimizing classification accuracy under varying data conditions. The Gazi Brains 2025 dataset plays a central role in this analysis, offering a unique and clinically representative sample set that enables a rigorous assessment of model performance in a real-world context. Complementarily, the inclusion of the BraTS 2021 dataset facilitates a cross-domain comparison, allowing for the evaluation of the consistency and transferability of augmentation effects across data sets with differing statistical and anatomical distributions. This comparative framework underscores the clinical and technical significance of augmentation robustness, particularly as demonstrated within the Gazi Brains 2025 dataset.

[5] https://www.kaggle.com/competitions/rsna-miccai-brain-tumor-radiogenomic-classification

The rest of this paper is structured as follows. Section 2 gives relevant works on generative and non-generative augmentation approaches for brain tumor classification. Also, multi-class classification of brain tumors is reviewed. Section 3 presents the used material and gives approaches used in binary and multiclass classifications of brain tumors. Section 4 firstly gives used datasets and then presents experimental results, and Section 5 gives a discussion and draws a conclusion as well as practical implications, limitations, and future works.

2 Related works

2.1 Generative-based augmentation approaches for brain tumor classification

In the evolving landscape of medical imaging, Generative Adversarial Networks (GANs) and diffusion models have emerged as transformative tools for brain tumor classification. These generative models enhance the quality and diversity of training datasets, thereby improving the accuracy and robustness of diagnostic algorithms.

GANs consist of a generator and a discriminator engaged in a dynamic contest, resulting in the creation of synthetic data that closely resembles real samples. In the context of brain tumor classification, GANs have been instrumental in augmenting MRI datasets. For instance, Yurtsever et al. [5] utilized GAN-based augmentation on the Gazi Brains 2020 dataset, leading to improved radiogenomic classification performance. Mahnoor et al. [11] explored the use of GANs to produce 64×64 pixel synthetic brain MRI images, aiming to augment the training data for a U-Net convolutional neural network. The inclusion of these GAN-generated images resulted in an increase in mean accuracy from 82% to 88%, demonstrating the efficacy of GAN-based augmentation in improving tumor detection models.

A study by Gab Allah et al. [12] investigated the application of Progressive Growing GANs (PGGANs) to synthesize high-resolution brain tumor MRIs across axial, sagittal, and coronal planes. The synthetic images were utilized to augment the training dataset, leading to enhanced performance of classification models. This approach highlighted the capability of PGGANs in generating detailed and diverse medical images. Yapici et al. [13] employed CycleGAN to generate synthetic images of glioma, meningioma, and pituitary tumors, which were then combined with original datasets to train a DenseNet121-based model. The incorporation of synthetic data improved classification accuracy across all tumor types, emphasizing the role of CycleGAN in enhancing deep learning models for medical diagnostics.

Alrashedy et al. [14] introduced BrainGAN, a framework that leverages GAN-generated images alongside deep learning models for brain MRI classification. The proposed method includes an automated validation process for evaluating the quality of synthetic images. Three deep learning models-CNN, MobileNetV2, and ResNet152V2-were trained using images generated by Vanilla GAN and Deep Convolutional GAN (DCGAN), and their performance was assessed on a test set of real brain MRI scans. Experimental results demonstrated that ResNet152V2 outperformed the other models. Nag et al. [15] integrated ResNet50 for feature extraction and GANs for data augmentation. Ten transfer learning models, including GoogLeNet and VGG-16, were evaluated for brain tumor classification using key performance metrics such as precision, recall, and F1-score, along with Hamming loss and the Matthews correlation coefficient for deeper analysis. To enhance model interpretability, Explainable AI (XAI) techniques, specifically Local Interpretable Model-Agnostic Explanations (LIME), were applied to ensure transparency in predictions. The study analyzed 7,023 MRI images, including cases of gliomas, meningiomas, pituitary tumors, and non-tumorous cases. Their proposed model, TumorGANet, demonstrated exceptional performance, achieving 99.53% accuracy, 100% precision and recall, 99% F1-score, and a Hamming loss of just 0.2%.

Brain tumor diagnosis relies on MRI for precise detection and treatment planning, but challenges such as data scarcity and the difficulty of acquiring specific modalities like T1 Contrast Enhanced (T1CE) hinder accurate segmentation. Kaliprasad and Indu [10] addressed these limitations by generating synthetic T1CE scans from T1 using an image-to-image translation framework (VarVit-GAN) and introducing Adaptive Random Patch Selection (ARPS) to improve segmentation by mitigating data sparsity. When integrated into the nnUNet model, their approach achieved a dice similarity coefficient (DSC) of 86.47%, demonstrating the effectiveness of synthetic data and advanced sampling techniques in enhancing brain tumor segmentation. Table 1 compares the brain tumor classification works using GAN-based augmentation.

Diffusion models, which learn to reverse a noise process to generate data, have shown remarkable efficacy in medical image analysis [17]. Wolleb et al. [18] demonstrated the application of diffusion models for implicit image segmentation ensembles, achieving detailed uncertainty maps and improved segmentation results. Further advancements include the development of multi-scale diffusion models with attention mechanisms, which have enhanced the precision of brain tumor segmentation [19]. Moreover, diffusion models have been employed for brain tumor inpainting, offering viable solutions for reconstructing affected brain regions and aiding in accurate diagnosis [20].

While both GANs and diffusion models aim to generate high-fidelity data, they differ in architecture and training dynamics. GANs are known for their rapid generation capabilities but can suffer from training instabilities and mode collapse. In contrast, diffusion models offer stable training and have been effective in generating diverse and realistic medical images, albeit often requiring longer training times. The choice between using GANs or diffusion models depends on specific application requirements, such as the need for data diversity, training stability, and computational resources. Table 2 compares brain MRI data augmentation for different tasks using diffusion-based approaches.

2.2 Non-generative-based augmentation approaches for brain tumor classification

In brain tumor classification, non-generative data augmentation techniques are essential for enhancing model performance, especially when dealing with limited datasets. These methods involve applying various transformations to existing MRI images to increase data diversity and improve the generalization capabilities of machine learning models.

Common non-generative augmentation techniques include (i) rotation and flipping: By rotating images at various angles and flipping them horizontally or vertically, models become invariant to orientation changes, aiding in accurate tumor localization. (ii) Scaling and zooming: Adjusting the size of images helps models recognize tumors at different resolutions and sizes, enhancing detection across varying tumor presentations. (iii) Elastic deformations: Applying slight elastic distortions to images simulates anatomical variability among patients, allowing models to generalize better to unseen data. (iv) Intensity variations: Modifying image brightness or contrast introduces variations in image intensity, enabling models to handle differences in MRI machine outputs and imaging conditions. Implementing these non-generative augmentation techniques has been shown to significantly improve the accuracy and robustness of deep learning models in brain tumor classification tasks [26, 27].

Anaya-Isaza and Mera-Jiménez [28] explored the impact of various conventional data augmentation techniques on brain tumor detection using ResNet50, incorporating a principal component analysis (PCA)-based approach as an augmentation approach alongside standard methods. Their proposed method, combined with transfer learning from ImageNet, achieved an F1 score of 92.34%, demonstrating its effectiveness and statistical significance ($p < 0.05$) compared to traditional augmentation techniques.

Table 1 Comparison of brain tumor classification works using GAN-based augmentation

Paper Name	Classifiers	Dataset	Data Augmentation Type	Evaluation Results
Classification of brain MRI tumor images based on deep learning PGGAN augmentation [12]	VGG19 + CNN, VGG19 + GRU, and VGG19 + Bi-GRU	Custom (Brain MRI Images)	PGGAN + Traditional (rotation, flip, mirror)	Accuracy: 98.54%
Development of brain tumor radiogenomic classification Using GAN-based augmentation [5]	MobileNetV2, Xception, EfficientNetV2S	Gazi Brains 2020, BraTS 2021, BR35H	StyleGANv2-ADA	Accuracy: up to 96.19%
TumorGANet: A transfer learning and generative adversarial network-based data augmentation model for brain tumor classification [15]	GoogleNet, DenseNet-121, and VGG-16	Kaggle Dataset	Basic GAN	Accuracy: 99.53%, Precision and recall: 100%, F1 score: 99%, and Hamming Loss: 0.2%
BrainGAN: brain MRI image generation and classification framework using GAN architectures and CNN models [14]	ResNet50	BraTS 2019	VarVit-GAN	DSC: 86.47%
Improving brain tumor classification with deep learning using synthetic data [13]	DenseNet1211-based CNN	Figshare brain dataset	CycleGAN	Accuracy: 99.47%
Exploring generative adversarial network-based augmentation of magnetic resonance brain tumor images [11]	Modified U-Net CNN	2000 transaxial two-dimensional brain MRI images	DC-GAN	Accuracy: 88%
Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images [16]	pre-trained DNN	Brain MRI	DCGAN	Accuracy: 91.70% Sensitivity: 90.16% Specificity: 95.58% Precision: 91.17% F1-score: 90.54%

Sreedhar et al. [29] introduced a novel approach that integrates Lipschitz-based data augmentation with modified AlexNet and network-based deep transfer learning for brain tumor classification. Evaluated on the BraTS 2020 and Figshare datasets, their proposed model demonstrated superior performance across multiple metrics, including sensitivity, specificity, accuracy, precision, F1-score, and MCC. Amena et al. [30] optimized VGG-16, VGG-19, and Inception-V3 architectures using the Aquila Optimizer (AQO) for brain tumor classification on augmented data with traditional approaches, achieving 98.95% accuracy with VGG-19 on a dataset of Glioma, Meningioma, and Pituitary tumors.

MRI brain images often contain noise, which reduces recognition accuracy and obscures critical details. To enhance image quality, noise and irrelevant regions must be minimized, particularly high-frequency noise, which is typically reduced through filtering techniques. The anisotropic diffusion filter is an effective method for autonomous noise suppression while preserving important image edges, preventing blurring and improving clarity for accurate analysis. Rahman and Islam [31] investigated the performance of MRI brain images after using an anisotropic diffusion filter.

Non-generative-based augmentation techniques are computationally efficient, requiring significantly less processing power compared to generative models like GANs or diffusion models. They also preserve the original data distribution, ensuring that no artificial artifacts distort medical images, which is crucial for accurate diagnosis. By exposing

Table 2 Comparison of brain MRI data augmentation using diffusion-based approaches

Paper Name	Task	Dataset	Data Augmentation Type	Evaluation Results
Conditional diffusion models for semantic 3D brain MRI synthesis [21]	Brain tumor segmentation	Clinical stereotactic radiosurgery dataset	Conditional diffusion model	Accuracy:66.75%
Discrepancy-based diffusion models for lesion detection in brain MRI [22]	Brain MRI lesion detection	BraTS 2020	Discrepancy distribution medical diffusion	Dice: 84.26%, MIoU: 88.58%
Diffusion transformer-augmented fMRI functional connectivity for enhanced autism spectrum disorder diagnosis [23]	fMRI data generation and augmentation	Collected rs-fMRI data	Brain-Net-Diffusion	Accuracy: 69.6%
Innovative multi-class segmentation for brain tumor MRI using noise diffusion probability models and enhancing tumor boundary recognition [24]	Multi-class segmentation for brain tumor MR	BraTS 2020	Diffusion probability models	Dice: 90.4%
Multi-modal MRI synthesis with conditional latent diffusion models for data augmentation in tumor segmentation [25]	Brain tumor segmentation	BraTS 2021	Slice-by-slice latent diffusion model	ET: 73.2% for real+synth

deep learning models to variations in the existing data, these techniques reduce overfitting and improve generalization. Additionally, they are easy to implement with most deep learning frameworks and work well with small datasets, making them a practical choice in medical imaging, where data availability is often limited.

However, non-generative augmentation has certain limitations. Since these methods only modify existing images rather than creating new ones, they provide limited data expansion, which may not be sufficient for deep learning models requiring diverse training samples. Overuse of simple transformations can also lead to redundancy, where augmented images do not contribute significant variability. Furthermore, in cases of severe data scarcity, these techniques may not provide enough diversity to train robust models. Unlike generative approaches, they do not enhance structural features or simulate realistic variations in tumor morphology, which can be crucial for detecting rare or atypical cases. As a result, while non-generative augmentation is a valuable tool for improving model performance, it may be insufficient for highly imbalanced datasets or cases requiring advanced feature enhancement.

2.3 Multi-classification of brain tumors

Accurate multi-class classification of brain tumors holds significant importance in the medical domain, as it directly contributes to precise diagnosis, effective treatment planning, and the advancement of personalized medicine approaches. Differentiating between various tumor types enables clinicians to tailor therapeutic strategies to individual patients, thereby improving prognostic outcomes and overall quality of care. Brain tumors come in various types, such as gliomas, meningiomas, pituitary tumors, and metastases, each with distinct characteristics and treatment requirements. Classifying these tumors into multiple categories rather than just differentiating between“benign" and “malignant" ensures that physicians can select the most appropriate treatment options tailored to the specific tumor type. This level of classification helps in understanding the tumor’s biology, predicting its behavior, and formulating effective treatment strategies, including surgery, radiation, or chemotherapy [32]. Additionally, accurate multiclass classification aids in the early detection of rare tumor types, which might otherwise be misdiagnosed when only considering a binary classification.

Furthermore, multiclass classification models can significantly improve clinical decision-making by providing a more comprehensive view of the patient’s condition [33]. They enable precise monitoring of tumor progression or response to treatment by distinguishing subtle variations between different tumor subtypes. This reduces the risk of misdiagnosis and treatment delays, which can be life-threatening, especially in aggressive tumor types. The ability to classify tumors into multiple categories is also crucial for clinical research, as it enhances the ability to study the relationship between tumor type, location, and patient outcomes.

Yadav et al. [34] propose a computer-aided diagnosis system for multi-class brain tumor classification using a modified VGG-16 architecture with transfer learning. The proposed approach incorporated a multi-view segregation of MRI images from axial, sagittal, and coronal planes, alongside an advanced preprocessing pipeline involving object-centric image extraction and contrast enhancement. The model is reported to achieve an accuracy of 98.5%, with an AUC of 99.0%, demonstrating competitive performance compared to existing methods. Zulfiqar et al. [35] proposed a transfer learning-based approach using fine-tuned EfficientNet models (B0–B4) to classify brain tumors into three categories: glioma, meningioma, and pituitary tumors, utilizing the publicly available Figshare CE-MRI dataset. Among the tested models, EfficientNetB2 achieved the best performance with a test accuracy of 98.86%, demonstrating strong generalization, low computational cost, and effective tumor localization through Grad-CAM visualizations. Kibriya et al. [36] introduced a lightweight 13-layer CNN model designed to automatically classify three types of brain tumors-glioma, pituitary, and meningioma-using MRI scans, eliminating the need for manual segmentation or handcrafted feature extraction. Their proposed model achieved 97.2% accuracy on a benchmark dataset and demonstrated strong generalization in a cross-dataset evaluation, proving its potential for practical, real-world clinical use in early tumor detection. Deependra et al. [37] presented a deep learning-based solution using a CNN enhanced with a multi-branch architecture and inception block, evaluated through five-fold cross-validation on the Br35H dataset, which includes images of gliomas, meningiomas, pituitary tumors, and healthy brains. Their proposed model achieved an outstanding accuracy of 99.30% with minimal variance (0.0005), significantly outperforming traditional classifiers like SVC, Decision Tree, and Logistic Regression, showcasing its potential for reliable and efficient brain tumor classification. Amin et al. [38] introduced a novel multi-class brain tumor classification model, ADE_DieT, which combines the DieT Transformer for feature extraction, PCA for dimensionality reduction, and the ADE algorithm for feature selection-achieving a 96.09% accuracy and demonstrating potential to support radiologists by increasing diagnostic speed and consistency in clinical practice.

Multi-class classification of brain tumors presents several significant challenges, both technical and clinical, that impact the reliability and applicability of automated diagnostic systems. One of the primary difficulties lies in intra-class similarity and inter-class variability. Tumors such as gliomas, meningiomas, and pituitary tumors often exhibit overlapping visual characteristics on MRI scans, making it difficult for machine learning models to distinguish among them with high precision. Furthermore, variations in tumor size, shape, location, and intensity across different patients increase the complexity of building a generalized model that performs consistently well across diverse cases. Class imbalance is another critical challenge-some tumor types appear more frequently in datasets than others, which can lead models to become biased toward the more dominant classes, thus reducing sensitivity and specificity for rarer tumor types. Another challenge is the limited availability of high-quality, annotated data. Medical imaging datasets are often small due to privacy restrictions, high labeling costs, and the expertise required for annotation. This scarcity makes it difficult to train deep learning models without overfitting.

Several architectures for brain tumor segmentation from FLAIR MRI have been proposed, including the EffUNet++ model by Yadav et al. [39] which integrates an EfficientNetB7 encoder, the DCRUNet++ model by Sonawane et al. [40] utilizing Depthwise Convolutional Residual modules, and the Mod-R2AU-Net by Yadav et al. [41] incorporating Recurrent Residual Convolutional Layers and Attention Gates. These approaches demonstrate high performance, with reported Dice coefficients of 0.9392, 0.9467, and an IoU of 0.9024 on BRATS 2019, respectively, outperforming previous benchmarks.

3 Material and methods

The overall methodology for evaluating the effect of augmented MRI data on classification accuracy is outlined in Fig. 1 and all codes related to methods are available at Github repo. The process begins with dataset preparation, which includes several key stages. Initially, raw DICOM records are collected, followed by the determination of segmentation labels. MRI data is then labeled and segmented using itk-SNAP software [42], and the dataset is organized by matching image-label pairs to ensure consistency between images and their corresponding labels.

The data preprocessing pipeline comprises multiple phases to standardize and prepare the MRI scans for model training. Initially, DICOM files are converted into the NIfTI format to facilitate efficient handling and interoperability. Subsequently, skull-stripping is performed utilizing region-of-interest (ROI) masks derived from the preceding segmentation process. Intensity normalization is applied to ensure consistency. Spatial normalization is then conducted, where Gazi Brains 2023 images are resampled to 256×256 pixels, while BRaTS 2021 images retain their native 240×240 resolution. For compatibility with 2D convolutional architectures, the preprocessed volumetric images are ultimately converted into 2D PNG slices for subsequent processing.

In the data augmentation phase, two advanced generative strategies are employed. The first utilizes a Guided Diffusion model to synthesize realistic variations of the original data, while the second leverages StyleGANv3 to generate high-fidelity synthetic images for augmentation purposes.

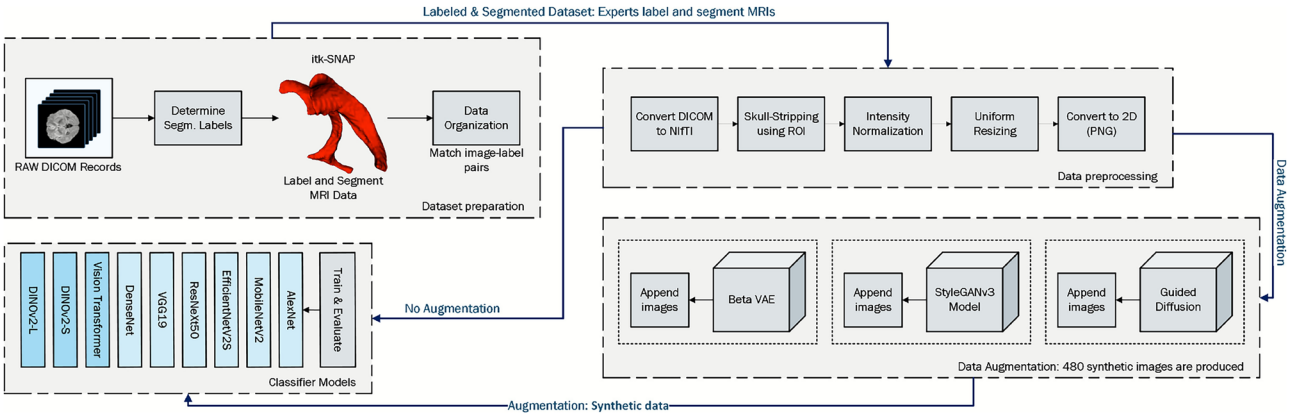


Fig. 1 Overview of the methodology for evaluating the effect of augmented MRI data on brain tumor classification accuracy. The process includes dataset preparation, data preprocessing, data augmentation using Guided-Diffusion, StyleGANv3, and Beta Variational Autoencoder, with training of eight different classifier models across three augmentation setups

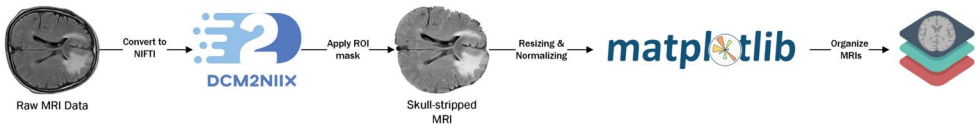


Fig. 2 Overview of the neuroimaging preprocessing pipeline for Gazi Brains 2023 dataset, including DICOM conversion, brain extraction, intensity normalization, resampling to 256×256, and 2D PNG generation

In the final phase, nine distinct deep learning classifiers are developed and trained: AlexNet, MobileNetV2, EfficientNetV2S, ResNeXt50, VGG19, DenseNet, Vision Transformer (ViT), DINOv2-S, and DINOv2-L. These models are evaluated under three experimental conditions: (1) without augmentation, (2) with augmentation via StyleGANv3, and (3) with augmentation via Guided Diffusion. This multi-scenario evaluation framework enables a comprehensive assessment of the impact of generative augmentation techniques on classification performance.

3.1 Preprocessing pipeline

The proposed preprocessing pipeline, illustrated in Fig. 2, encompasses a series of standardized steps to ensure consistent and high-quality input data for model training. First, raw DICOM images are converted to compressed NIfTI format, preserving metadata while enabling efficient processing. Next, skull-stripping is applied to isolate brain tissue, followed by intensity normalization to mitigate scanner-induced variability. Spatial normalization is then performed, where Gazi Brains 2023 scans are resampled to 256×256 voxels, while BRaTS 2021 scans retain their original 240×240 resolution. Finally, the 3D volumes are sliced into 2D PNG images and organized by diagnostic labels for supervised learning.

Converting from DICOM to NIfTI data format

The raw DICOM files containing medical imaging data and metadata (e.g., patient information, scan parameters) are processed using dcm2niix tool[6]. This tool extracts the pixel data and relevant metadata, then converts each scan series into a compressed NIfTI (.nii.gz) format. NIfTI is preferred for its compatibility with neuroimaging tools and efficient storage.

Skull-Stripping using ROI Information

Non-brain tissues (e.g., skull, scalp, meninges) are removed using the provided ROI annotation data that experts provide with Gazi Brains 2025 dataset. A region of interest (ROI) is manually defined to isolate brain matter, improving segmentation accuracy and reducing computational noise from irrelevant structures.

Intensity Normalization and Uniform Resizing

To account for variations in scanner protocols and acquisition settings, voxel intensities are normalized using min-max scaling. This ensures consistent input ranges across scans, preventing bias due to contrast or brightness differences. With uniform resizing, MRI slices are resampled to a fixed resolution. This standardization ensures uniform input dimensions for deep learning models, avoiding shape mismatches during batch processing.

Converting Medical Image to the 2D

Each 3D volume (NIfTI) is sliced along the axial, sagittal, or coronal plane to generate 2D images. These slices are saved as PNG files and organized into labeled directories (healthy/tumor) for supervised learning. This step facilitates compatibility with deep learning architectures.

[6] https://github.com/rordenlab/dcm2niix

3.2 Data augmentation strategies

3.2.1 StyleGAN3-T augmentation

StyleGAN3-T is a generative model designed for high-quality image synthesis with a focus on translation invariance and improved texture generation (see Algorithm 1). It is a variant of StyleGAN3, which builds on the capabilities of StyleGAN2 but introduces architectural changes to refine image consistency and spatial coherence. StyleGAN3-T, in particular, is optimized for texture-heavy applications, making it suitable for tasks requiring fine-grained details [43]. Unlike previous versions, StyleGAN3-T removes artificial positional encoding in the synthesis network, allowing images to maintain realism without fixed pixel positions. This prevents unwanted artifacts and ensures smooth transformations across generated images. StyleGAN3-T introduces a continuous signal processing approach to improve the way textures are learned and generated. This makes it more effective at replicating fine patterns, reducing distortions seen in StyleGAN2. By using advanced wavelet-based transformations and adaptive learning rates, StyleGAN3-T achieves more natural-looking outputs.

Initialize StyleGAN3-T Model
Load pre-trained generator G and discriminator D
Input Random Noise ▷ Start data generation by sampling random latent codes, which drive diversity in the output images.
for each sample i **do**
    Sample latent vector $z_i \sim P(z)$
    Map z_i to intermediate latent space
end for
Generate Synthetic Images ▷ Use the generator to produce images from the intermediate latent vectors.
for each z_i **do**
    Apply learned style transformations
    Generate synthetic image $x_i' = G(z_i)$
end for
Post-processing ▷ Optionally clean up or refine the images before using them in training.
if image requires enhancement **then**
    Apply normalization, filtering, or other refinements
end if
Augmentation Pipeline ▷ Further augment the synthetic images with traditional data augmentation techniques.
for each synthetic image x_i' **do**
    Apply transformations (scaling, flipping, rotation) if needed
    Ensure consistency and realism in generated data
end for
Use Augmented Data ▷ Combine real and synthetic data for training.
Add generated samples to dataset
Train downstream model using real + synthetic data
Return enhanced dataset with augmented images

Algorithm 1 StyleGAN3-T based augmentation

3.2.2 Guided diffusion-based augmentation

Guided Diffusion augmentation (see Algorithm 2) is a data augmentation technique that leverages diffusion models to generate high-quality synthetic samples while maintaining control over the output distribution. It is based on denoising diffusion probabilistic models (DDPMs), which iteratively refine noisy inputs to generate realistic images. By incorporating guidance mechanisms, this approach enhances the generated samples' fidelity and diversity, making them more useful for training machine learning models [44].

Initialize Parameters
Set number of timesteps T
Define variance schedule $\{\beta_t\}$ for noise addition
Initialize neural network θ for denoising
Forward Diffusion Process ▷ Simulate the diffusion process: starting from a clean image, gradually add noise over T steps.
for $t = 1$ to T **do**
    Sample noise $\varepsilon_t \sim \mathcal{N}(0, I)$
    Apply noise to input image
    $x_t = \sqrt{1-\beta_t} \cdot x_{t-1} + \sqrt{\beta_t} \cdot \varepsilon_t$
end for
Reverse Diffusion Process (Denoising) ▷ Start from pure noise and iteratively denoise toward the clean image.
for $t = T$ to 1 **do**
    Predict noise using neural network: $\varepsilon_\theta(x_t, t)$
    Compute mean $\mu_\theta(x_t, t)$ for denoising
    $\mu_\theta(x_t, t) = \frac{x_t - \sqrt{\beta_t} \cdot \varepsilon_\theta(x_t, t)}{\sqrt{1-\beta_t}}$
    Sample x_{t-1} using learned parameters
    $x_{t-1} = \mu_\theta(x_t, t) + \sigma_t \cdot \varepsilon$
end for
Apply Guidance ▷ Use external guidance (classifier-based or classifier-free) to steer generation toward desired semantics.
if Using Classifier Guidance **then**
    Compute log-probability gradient $\nabla_x \log p(y|x_t)$
    Adjust sampling with guidance weight w
    $x_t \leftarrow x_t + w \cdot \nabla_x \log p(y|x_t)$
end if
if Using Classifier-Free Guidance **then**
    Compute guided noise estimate
    $\hat{\varepsilon}_\theta(x_t, y) = (1 + w) \cdot \varepsilon_\theta(x_t, y) - w \cdot \varepsilon_\theta(x_t)$
    Adjust denoising step accordingly
end if
Generate Augmented Sample ▷ Final output after reverse diffusion; this is the new, guided image.
Return final synthesized sample x_0
Ensure output meets augmentation criteria
Repeat for multiple samples to enhance dataset diversity.

Algorithm 2 Guided diffusion augmentation

3.2.3 Beta variational autoencoder (BetaVAE)

Beta Variational Autoencoder (ß-VAE, see Algorithm 3) is a modification of the standard variational autoencoder designed to learn interpretable, disentangled representations of data in a fully unsupervised manner. By introducing a hyperparameter ß that balances reconstruction accuracy with constraints on the latent space, ß-VAE encourages the model to capture independent generative factors of variation in the data. This approach improves the stability and interpretability of learned features, making it particularly useful for tasks requiring factor-level understanding without labeled data [45].

Initialize encoder $q_\phi(\mathbf{z}|\mathbf{x})$ and decoder $p_\theta(\mathbf{x}|\mathbf{z})$
Set hyperparameter $\beta > 1$ ▷ Controls disentanglement vs reconstruction trade-off
for each training batch $\mathbf{x}$ **do**
    Encode: $\mu, \sigma = \text{encoder}(\mathbf{x})$
    Sample: $\mathbf{z} \sim \mathcal{N}(\mu, \sigma^2)$ ▷ Reparameterization trick
    Decode: $\hat{\mathbf{x}} = \text{decoder}(\mathbf{z})$
    Compute reconstruction loss: $\mathcal{L}_{\text{recon}} = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$
    Compute KL divergence: $\mathcal{L}_{\text{KL}} = D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})\|\mathcal{N}(0, I))$
    Total loss: $\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta \cdot \mathcal{L}_{\text{KL}}$ ▷ β strengthens latent constraints
end for
Return trained encoder and decoder with disentangled latent space

Algorithm 3 β-VAE Training

3.3 Classification methods

3.3.1 Vision transformers

ViT-Base (Vision Transformer-Base)

It is a deep learning model architecture designed for image classification tasks, based on the Transformer architecture, which was originally developed for natural language processing. Unlike traditional CNNs, ViT-Base splits an image into fixed-size patches, linearly embeds them, and processes the sequence of patch embeddings using self-attention mechanisms to capture spatial relationships within the image. This approach allows ViT to scale effectively with large datasets and provides flexibility in learning complex patterns. The ViT-Base model, which typically consists of 12 layers and 768 hidden units, has demonstrated strong performance on large-scale benchmarks like ImageNet, surpassing traditional CNNs in some cases, especially when pre-trained on vast datasets and fine-tuned for specific tasks. Its success highlights the potential of Transformer-based models in computer vision [46].

DINOv2 (Distillation with No Labels version 2)

It is a self-supervised learning method for vision models that leverages knowledge distillation to learn robust visual representations without the need for labeled data. The model operates by encouraging agreement between student and teacher networks on image representations, with the teacher network being a slowly updated version of the student. DINOv2 comes in different variants, with the “S" (small) version using 384-dimensional embeddings and the “L" (large) version using 1024-dimensional embeddings. The larger embedding size in the “L" variant allows for a richer, more expressive representation, which can improve performance on a variety of downstream tasks, especially when dealing with more complex visual data. These variants allow for scalability in model size and performance, providing flexibility depending on computational resources and the specific needs of the task [47]. DINOv2’s effectiveness has been shown in a range of computer vision tasks, making it a powerful tool for learning high-quality, unsupervised visual representations.

3.3.2 Convolutional networks

This study evaluates a range of state-of-the-art CNN-based architectures for MRI-based brain anomaly classification. The selected models-AlexNet, MobileNetV2, EfficientNetV2S, ResNeXt50, VGG19, and DenseNet-offer a balance between depth, efficiency, and representational power. These architectures were tested under consistent preprocessing and augmentation settings to enable fair performance comparison. A concise overview of each model is provided below.

AlexNet

AlexNet is a pioneering deep CNN architecture that revolutionized the field of computer vision. It consists of eight layers: five convolutional layers followed by three fully connected layers. AlexNet uses ReLU (Rectified Linear Unit) activation functions, which helped overcome the vanishing gradient problem and significantly accelerated training. Additionally, it introduced techniques like data augmentation and dropout to reduce overfitting [48].

MobileNetV2

MobileNetV2 is a lightweight and efficient convolutional neural network (CNN) architecture designed for mobile and edge devices, offering a good balance between performance and computational efficiency [49]. This architecture utilizes depthwise separable convolutions, which reduce the computational burden by applying separate convolutions to individual channels rather than all input-output channel pairs. The inverted residual block improves efficiency by first expanding the feature maps with a lightweight depthwise convolution, followed by a bottleneck structure that reduces dimensions, and finally applying a pointwise convolution.

EfficientNetV2S

Table 3 Pathological MRI records and descriptions in the Gazi Brains 2025 dataset

Anomaly	Description	Slice count
Cavernoma	Abnormal cluster of blood vessels in brain tissue	133
High-Grade Glioma	Malignant CNS tumors (WHO Grades 3–4)	499
Low-Grade Glioma	Slow-growing CNS tumors (WHO Grades 1–2)	118
Metastasis	Secondary brain tumors from systemic cancers	123
Meningioma	Extra-axial tumors arising from meninges	576
Postoperative	Post-surgical imaging findings	127
Schwannoma	Benign nerve sheath tumors	105

Table 4 Segmentation labels of the Gazi Brains 2025 dataset

Label	Name	Label	Name
0	Empty Label	11	Cavum Septum Pellucidum
1	ROI – Region of Interest	12	Hemorrhage
2	Eye	13	Acute Infarction
3	Optic Nerve	14	Cystic Area
4	Lateral Ventricle	15	Arachnoid Cyst
5	Third Ventricle	16	Cavernoma
6	Ischemic Gliotic Areas	17	Encephalomalacic Area
7	Non-Enhancing Tumor Portion	18	Subdural Space
8	Peritumoral Edema Area	19	Operation Cavity
9	Contrast-Enhancing Tumor Portion	20	Demyelination
10	Tumor Necrosis	21	Air
		22	Calcification

EfficientNetV2S is a variant of the EfficientNetV2 family, which is designed to optimize the trade-off between model accuracy and computational efficiency [50]. It utilizes a combination of neural architecture search (NAS) and a novel compound scaling method to balance network depth, width, and resolution. The architecture incorporates lightweight depthwise convolutions and incorporates techniques like training with progressive resizing, which gradually increases image resolution during training for improved performance.

ResNeXt50

ResNeXt50 is a CNN architecture that builds on the ResNet (Residual Network) framework, incorporating the concept of "cardinality," which refers to the number of parallel paths or groups within a layer. This design is aimed at improving the representational power of the network while maintaining efficiency. The ResNeXt50 model consists of 50 layers and is known for using grouped convolutions, where each group performs a convolution independently before combining the results [51]. This structure allows the network to learn more diverse features, leading to better performance with fewer parameters compared to traditional architectures.

VGG19

VGG19 is a CNN architecture belonging to the VGG family, renowned for its uniform design and effectiveness in image classification tasks. It consists of 19 layers, including 16 convolutional layers and 3 fully connected layers, with a simple and uniform structure that uses small 3x3 convolutional filters with a stride of 1 and 2x2 max-pooling layers with a stride of 2. VGG19's key strength lies in its depth, which allows the model to capture hierarchical features from images [52].

DenseNet (Densely Connected Convolutional Network)

It is a deep learning architecture that introduces a novel connectivity pattern, where each layer is connected to every other layer in a feed-forward manner. In contrast to traditional CNNs, where each layer receives input from only the previous layer, DenseNet ensures that every layer receives feature maps from all preceding layers [53]. This dense connectivity allows for efficient feature reuse, reducing the need for redundant feature maps and alleviating the vanishing gradient problem by providing more direct paths for gradients during backpropagation. DenseNet architectures are typically more parameter-efficient than traditional CNNs, as they require fewer parameters to achieve high accuracy.

4 Experimental results

Experimental studies of the approaches on data sets and the obtained outputs are presented in this section.

4.1 BraTS 2021 dataset

The BraTS 2021 (Brain Tumor Segmentation) dataset [54] is a widely used benchmark dataset in the field of medical image analysis, particularly for the development and evaluation of deep learning models for brain tumor detection, segmentation, and classification. It was released as part of the RSNA-ASNR-MICCAI BraTS 2021 challenge, which aimed to foster research on automatic segmentation of brain tumors and their subregions from multimodal MRI scans. The dataset provides expertly annotated scans collected from more than 20 medical institutions, ensuring diversity in imaging protocols and patient demographics, which is essential for building robust and generalizable AI models.

Each case in the BraTS 2021 dataset includes four 3D MRI modalities: T1-weighted (T1), contrast-enhanced T1-weighted (T1Gd), T2-weighted (T2), and Fluid-Attenuated Inversion Recovery (FLAIR). These modalities are co-registered, skull-stripped, and resampled to a uniform

size of 240×240×155 voxels. The dataset focuses on gliomas, including high-grade glioblastomas (GBM) and lower-grade gliomas (LGG), which differ significantly in biological behavior and prognosis. All scans are accompanied by detailed segmentation labels that identify three tumor subregions: the enhancing tumor (ET), the peritumoral edema (ED), and the necrotic/non-enhancing tumor core (NCR/NET). These labels are further grouped into clinically relevant combinations such as whole tumor (WT), tumor core (TC), and enhancing tumor (ET).

The BraTS 2021 training set contains 1251 annotated cases, while the validation and test sets are held out for challenge submissions. The segmentation labels are used to assess model performance in tasks like tumor delineation, volume measurement, and progression tracking. The dataset also supports research in radiogenomic classification, making it valuable for both imaging and genomics-based studies. Due to its rich annotations, standardized format, and clinical relevance, BraTS 2021 has become a cornerstone in the development of automated brain tumor diagnosis systems, helping researchers advance the field of AI in healthcare.

4.2 Gazi Brains 2025 dataset

Gazi Brains 2025 dataset represents an expanded and enhanced version of the original Gazi Brains 2020 dataset[7], developed through a collaborative effort among medical experts at Gazi University Faculty of Medicine. Although the earlier dataset focused exclusively on high-grade glioma, this new iteration broadens the scope to include seven distinct brain anomalies, such as low-grade glioma, arteriovenous malformation, meningioma, and cavernoma, along with their corresponding segmentation masks. The dataset's pathological class balance, detailed in Table 3, reflects clinical incidence rates for primary brain tumors.

The dataset consists of MRI scans from 500 patients, including 93 scans classified as normal (healthy) and 407 scans containing pathological findings. Each pathological case is meticulously annotated by radiologists using ITK-SNAP[8] (versions 3.6.0 and 3.8.0), with regions of interest (ROIs) carefully delineated to ensure accurate tumor localization and segmentation. This expansion not only increases the dataset's diversity but also enhances its utility for research by providing detailed anatomical and pathological labels.

In terms of composition, the dataset is divided into two primary classes: Pathologic, which combines MRI scans of seven different anomaly types (e.g., HGG, LGG, AVM, meningioma), and Normal, which comprises scans without any pathological findings. The pathologic subset includes data from 407 patients, yielding 1,681 slices after preprocessing, while the normal subset consists of 93 patients, resulting in 997 preprocessed slices. For a detailed breakdown of label distributions, please refer to Table 4.

Key advancements over the original Gazi Brains 2020 dataset include greater diversity (seven anomaly types compared to the previous HGG-only focus), a larger scale (500 patients versus a smaller initial cohort), and improved annotations (expert-reviewed segmentations with precise ROIs). These enhancements make Gazi Brains 2025 a more comprehensive and robust resource for brain lesion analysis and computational research.

4.3 Experimental setup

An overview of the experimental setup is provided in this section. StyleGANv3 and Guided-Diffusion models are trained on a system with an Intel Xeon Gold 6336Y processor, 512GB RAM, and four NVIDIA A40 GPUs with 48GB VRAM each, running Ubuntu 20.04 LTS. Other deep learning and transformer models are trained on a system with an Intel i7-14700K processor, 32GB RAM, and one NVIDIA RTX 3090 GPU with 24GB VRAM, using WSL2 on Windows.

All code is implemented in PyTorch. The RMSProp optimizer is used with a learning rate of 1×10^{-4}. The Gazi Brains 2025 dataset is annotated using ITK-SNAP versions 3.6.0 and 3.8.0. DICOM to PNG conversion is performed using dcm2niix. Python 3.11.11 is used for development.

All architectures used in this study serve directly as baseline methods without any modification, ensuring fair performance comparison. A consistent 70%–15%–15% split is applied for training, validation, and testing across all experiments. The dataset distributions are as follows: Gazi Brains 2025 contains 997 normal and 1681 anomaly samples, while BraTS21 includes 2111 normal and 2572 anomaly samples. Each configuration in the ablation study is trained independently five times using fixed random seeds to ensure statistical stability and reproducibility. Performance differences therefore reflect only the model characteristics and augmentation strategies, supported by a uniform data split and controlled experimental conditions.

4.4 Performance metrics

A comprehensive set of performance metrics is employed to evaluate the effectiveness and reliability of the proposed models across both discriminative and generative tasks. These metrics were selected to capture not only predictive accuracy but also the quality of learned representations and the consistency of generated outputs.

[7] https://www.kaggle.com/datasets/gazibrains2020/gazi-brains-2020

[8] http://www.itksnap.org

Table 5 Binary classification performance on BraTS 2021 dataset

Model	Vanilla		StyleGANv3		Guided-Diffusion	
	Loss	Acc. (%)	Loss (Δ %)	Acc. (Δ %)	Loss (Δ %)	Acc. (Δ %)
AlexNet	0.4973	77.1	0.5201 (+4.57)	75.98 (−1.46)	0.5934 (+19.36)	74.50 (−3.37)
MobileNetV2	0.5574	75.25	0.5110 (−8.31)	79.40 (+5.52)	0.5493 (−1.45)	75.56 (+0.41)
EfficientNetV2S	0.5323	76.81	**0.5257 (−1.24)**	**74.28 (−3.30)**	**0.4603 (−13.51)**	**80.98 (+5.46)**
ResNeXt50	0.5917	72.83	**0.4420 (−25.34)**	**78.22 (+7.40)**	**0.6099 (+3.07)**	**77.41 (+6.28)**
VGG19	0.5731	68.71	0.5269 (−8.06)	73.75 (+7.35)	0.5444 (−4.99)	73.32 (+6.72)
DenseNet	0.4640	80.94	0.5960 (+28.47)	80.97 (+0.04)	0.4680 (+0.86)	81.77 (+1.02)
ViT	0.6488	63.02	0.6308 (−2.77)	67.19 (+6.59)	0.6160 (−4.78)	65.92 (+4.60)
DINOv2-S	0.5182	75.82	0.4829 (−6.83)	77.30 (+1.96)	0.4762 (−8.11)	76.09 (+0.35)
DINOv2-L	0.4775	78.52	0.4313 (−9.69)	76.64 (−2.39)	0.4955 (+3.78)	79.00 (+0.61)

Bold texts indicate the best values

Benchmarking model efficacy for brain tumor detection using StyleGANv3 and guided diffusion augmentation techniques, compared to vanilla models. The performance is measured using standard evaluation metrics, namely loss and accuracy, with percentage changes calculated to assess the effectiveness of augmentation

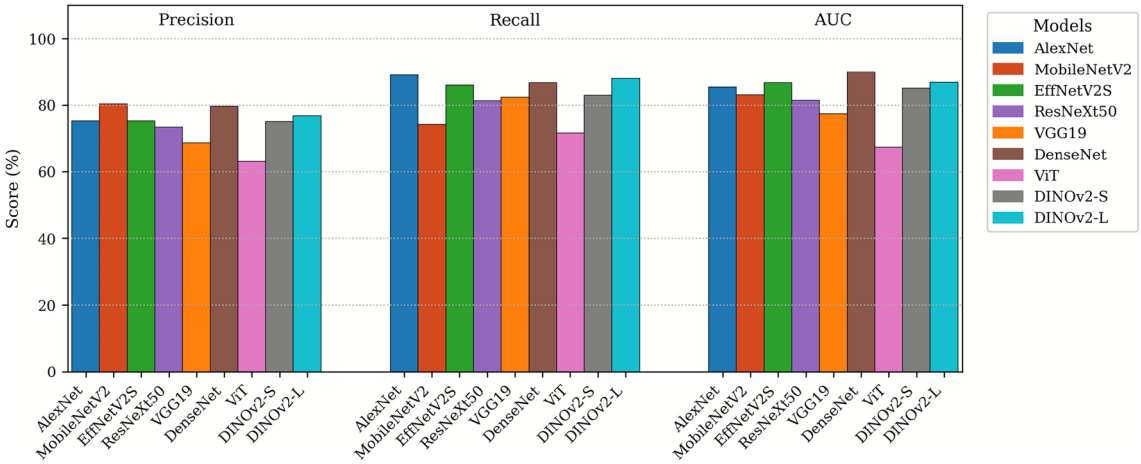


Fig. 3 Performance metrics of vanilla models on the BraTS 2021 dataset

For binary classification tasks, Binary Cross-Entropy (BCE) is utilized as the loss function, measuring the divergence between predicted probabilities and true labels as seen in Eq. 1.

$$\text{BCE} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right] \tag{1}$$

where y_i represents the true label, p_i denotes the predicted probability for class 1, and N is the number of samples.

Standard evaluation metrics such as precision, recall, and F1-score are used for the classification task. Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is evaluated to measure the trade-off between TPR and FPR. For multi-class classification tasks, Categorical Cross-Entropy (CCE) is used as the loss function as given in Eq. 2.

$$\text{CCE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(p_{i,c}) \tag{2}$$

where C is the number of classes, $y_{i,c}$ is a binary indicator of class membership, and $p_{i,c}$ is the predicted probability for class c.

Macro-averaged Precision, Recall, and F1-score are computed across all classes to provide an overall measure of classification quality.

For generative models such as StyleGAN and diffusion-based architectures, standard metrics like Fréchet Inception Distance (FID), Inception Score (IS), and Equivariance Scores (EQ-T, EQ-R) are utilized. FID is used to compare feature distributions between generated and real images. EQ-T and EQ-R measure temporal and rotational consistency in generated outputs, respectively. These metrics help evaluate not only visual fidelity but also the structural and semantic consistency of the generated data. IS evaluates the diversity and quality of generated samples, defined as in Eq. 3.

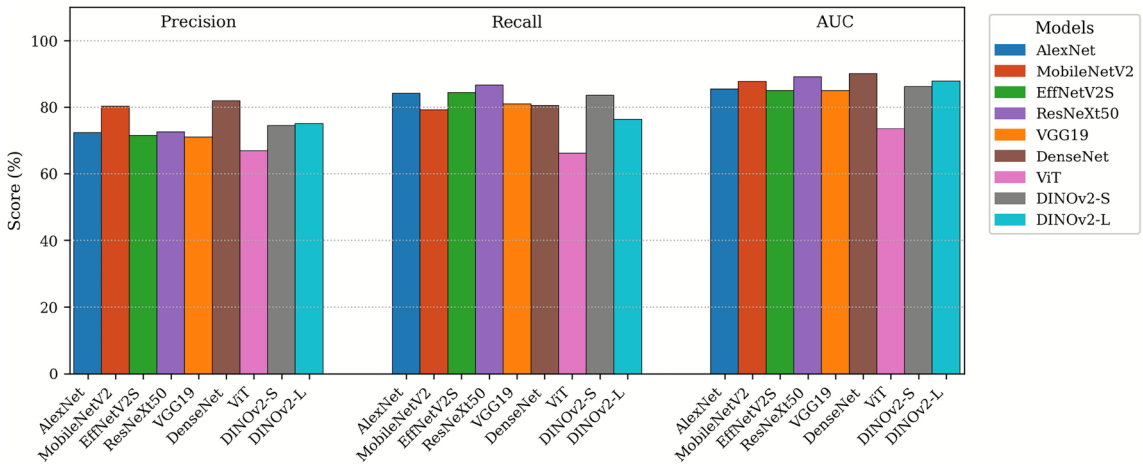


Fig. 4 Performance metrics of models augmented with StyleGANv3 on the BraTS 2021 dataset

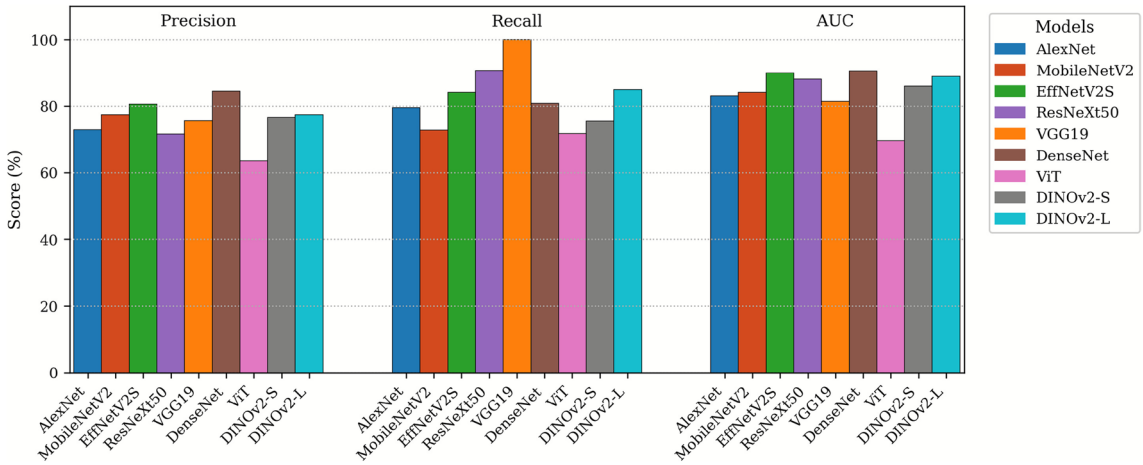


Fig. 5 Performance metrics of models augmented with Guided-Diffusion on the BraTS 2021 dataset

Table 6 Binary classification performance on the Gazi Brains 2025 dataset

Model	Vanilla		StyleGANv3		Guided-Diffusion	
	Loss	Acc. (%)	Loss (Δ %)	Acc. (Δ %)	Loss (Δ %)	Acc. (Δ %)
AlexNet	0.6530	63.77	0.3399 (−47.95%)	0.8405 (+31.75%)	0.6551 (+0.31%)	0.5884 (−7.73%)
MobileNetV2	0.3611	85.61	0.3228 (−10.61%)	0.8707 (+1.71%)	0.2815 (−22.06%)	0.8815 (+2.97%)
EfficientNetV2S	0.3314	84.86	0.2183 (−34.09%)	0.9181 (+8.20%)	0.3128 (−5.62%)	0.8664 (+2.10%)
ResNeXt50	0.3688	84.37	0.3471 (−5.93%)	0.8966 (+6.25%)	0.3076 (−16.60%)	0.8879 (+5.21%)
VGG19	0.6597	63.03	0.4801 (−27.27%)	0.7974 (+26.49%)	0.6898 (+4.57%)	0.5776 (−8.37%)
DenseNet	0.3200	86.10	**0.2456 (−23.13%)**	**0.9138 (+6.12%)**	**0.2440 (−23.75%)**	**0.9116 (+5.88%)**
ViT	0.5135	75.68	0.3978 (−22.59%)	0.8276 (+9.37%)	0.4805 (−6.43%)	0.7866 (+3.94%)
DINOv2-S	0.3774	81.14	0.3475 (−7.94%)	0.8319 (+2.53%)	0.3607 (−4.43%)	0.8276 (+2.00%)
DINOv2-L	0.2778	87.59	0.2900 (+4.33%)	0.8685 (−0.84%)	0.3136 (+13.01%)	0.8448 (−3.55%)

Bold texts indicate the best values

Model performance is evaluated for brain tumor detection using the original Gazi Brains 2025 dataset as well as its augmented counterparts generated via StyleGANv3 and Guided-Diffusion. Standard classification metrics, including loss and accuracy, are used. Percentage changes relative to the vanilla model are indicated in parentheses

$$\mathrm{IS} = \exp\left(\mathbb{E}_{x\sim p_g}\left[\mathrm{KL}(p(y \mid x)\|p(y))\right]\right) \tag{3}$$

where $p(y \mid x)$ is the conditional class distribution, and $p(y)$ is the marginal distribution.

In addition to FID and IS, the precision and recall of generated samples with respect to the real data distribution were also calculated, including spatial FID (sFID) for perceptual quality assessment.

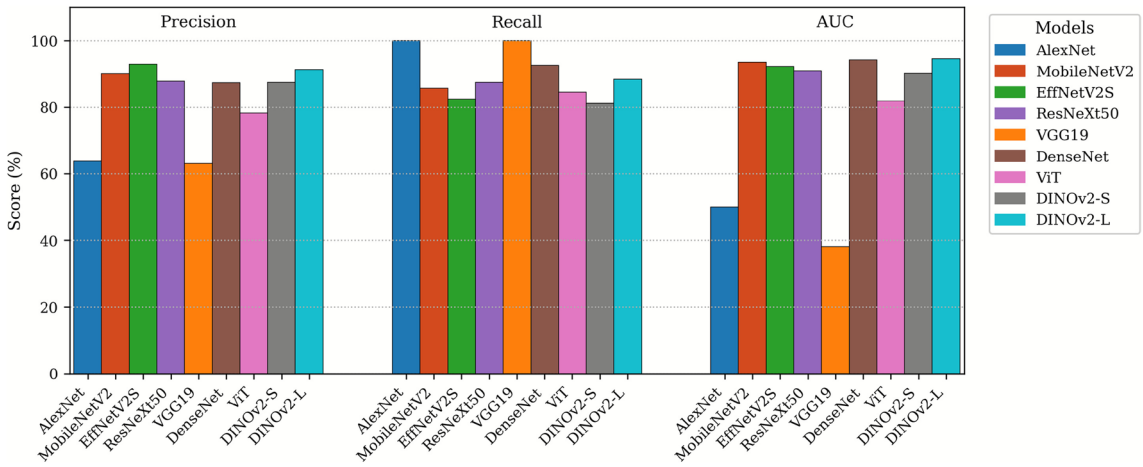


Fig. 6 Performance metrics of vanilla models on the Gazi Brains 2025 dataset

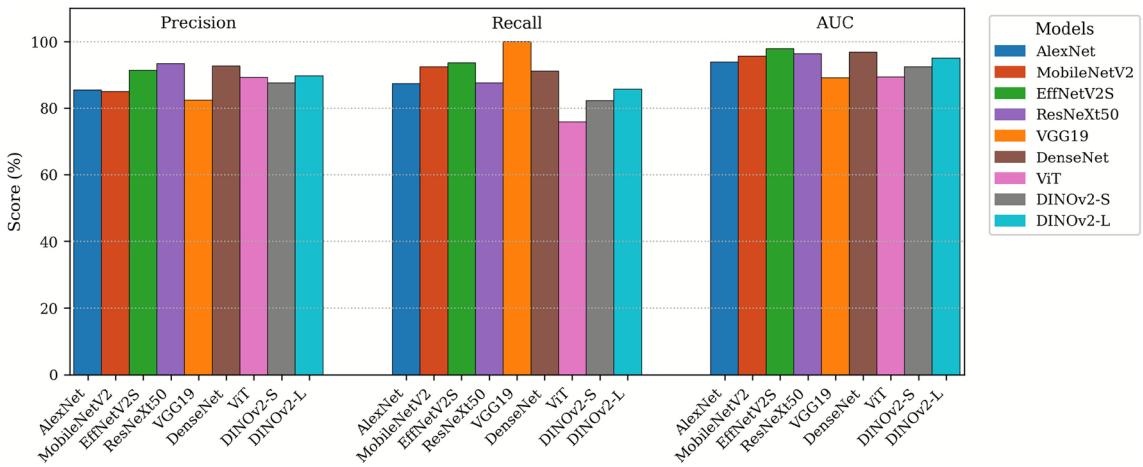


Fig. 7 Performance metrics of models augmented with StyleGANv3 on the Gazi Brains 2025 dataset

4.5 Binary classification results for anomaly detection

In this study, several deep learning models are employed to evaluate the effect of augmentation on binary classification performance. The augmentation strategies used include the generation of synthetic data through StyleGANv3 and guided diffusion, both aimed at enhancing model generalization and improving accuracy. To assess the efficacy of these augmentation techniques, the models are evaluated using loss and accuracy metrics.

The improvement criterion used in this study involves the calculation of percentage changes in both loss and accuracy, comparing the augmented models to their respective vanilla counterparts. For clarity, the percentage change (PC) is computed as in Eq. 4.

$$\mathrm{PC} = \left(\frac{\mathrm{NV} - \mathrm{OV}}{\mathrm{OV}} \right) \times 100 \tag{4}$$

Here, *New value-NV* refers to the performance metric (loss or accuracy) of the augmented model (either StyleGANv3 or guided diffusion), and *Old value-OV* corresponds to the baseline (vanilla model performance). Positive percentage changes in accuracy reflect an improvement, while negative changes indicate a decline. Similarly, positive changes in loss signify a worsening of performance, and negative changes indicate an improvement.

The results of the classifier models on BraTS 2021 dataset that take advantage of both StyleGANv3-based and diffusion-based augmentation are given in Table 5. The results show that different models benefited to varying degrees from the augmentation techniques. Some models, such as EfficientNetV2S and Resnext50_32x4d, experienced significant improvements in accuracy, with EfficientNetV2S showing a notable gain of +5.46% in accuracy when augmented with guided diffusion. On the other hand, models like AlexNet and VGG19 demonstrated less pronounced improvements in accuracy and sometimes even experienced a slight decrease in performance, suggesting that not all

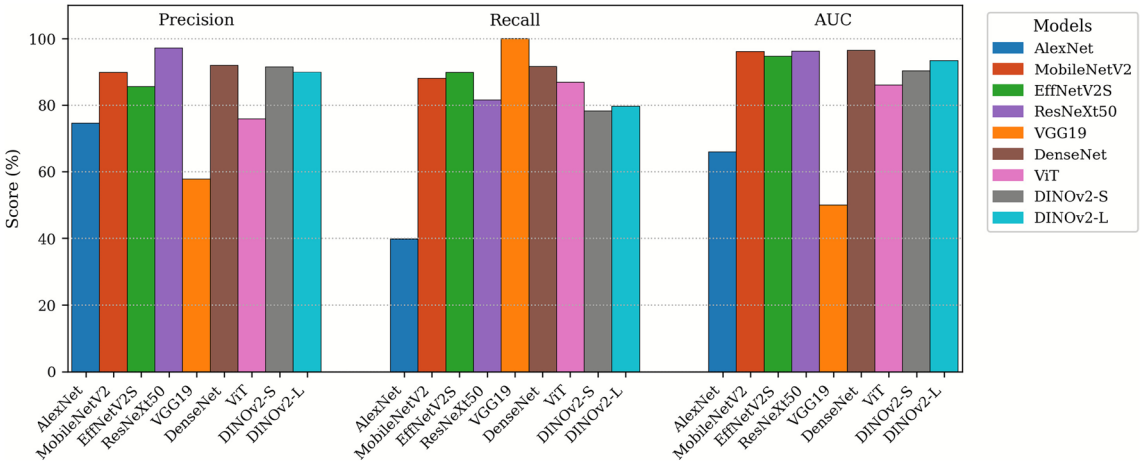


Fig. 8 Performance metrics of models augmented with Guided-Diffusion on the Gazi Brains 2025 dataset

Table 7 Ablation study on synthetic data augmentation using BetaVAE model

BetaVAE Augmentation						
Training set	Gazi Brains 2025			BRaTS21		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
vanilla	84.32	88.20	86.10	75.71	77.27	78.22
vanilla + 200 synthetic	89.27 ± 2.23	87.87 ± 1.71	86.61 ± 1.12	82.32 ± 3.68	80.11 ± 3.87	80.63 ± 3.18
vanilla + 300 synthetic	87.48 ± 4.11	88.54 ± 2.99	86.07 ± 1.66	80.52 ± 5.68	80.13 ± 7.55	79.92 ± 4.51
vanilla + 480 synthetic	88.01 ± 2.17	87.86 ± 2.88	87.37 ± 1.31	79.58 ± 3.22	81.46 ± 1.31	80.21 ± 2.88

Best-performing architectures (DenseNet for Gazi Brains 2025, ResNeXt50 for BRaTS21) were fine-tuned with 200, 300, and 480 synthetic images. Results show mean ± std. dev. across 5 independent runs

models benefit equally from augmentation. Figures 3, 4, 5 also show performance metrics of vanilla, StyleGANv3-based, and diffusion-based models on the BraTS 2021 dataset.

To investigate the effectiveness of data augmentation strategies in brain tumor classification, several deep learning models are evaluated on the Gazi Brains 2025 dataset under three settings: vanilla (original), StyleGANv3-augmented, and Guided-Diffusion-augmented. For each setting, both loss and accuracy metrics are considered, and percentage changes are computed to quantify the impact of augmentation. The results of the classifier models that take advantage of both GAN-based and diffusion-based augmentation for Gazi Brains 2025 dataset are given in Table 6.

Table 6 provides a detailed evaluation of how GAN-based and diffusion-based data augmentation techniques influence the performance of various deep learning classifiers on the Gazi Brains 2025 dataset. Among the tested architectures, DenseNet yielded the highest classification accuracy, reaching 91.38% with StyleGANv3-augmented data and 91.16% with Guided-Diffusion. These results correspond to absolute improvements of +5.28% and +5.88%, respectively, when compared to the baseline performance without augmentation. MobileNetV2 also showed notable gains-particularly with Guided-Diffusion (+2.97%)-achieving an overall accuracy of 88.15%. Similar positive trends were observed for EfficientNetV2S and ResNeXt50, both of which benefited considerably from StyleGANv3-enhanced data, with improvements of +2.10% and +5.27%, respectively.

On the other hand, models such as DINOv2-L and VGG19 experienced performance degradation following augmentation, highlighting that larger or transformer-based models may require more refined augmentation strategies to prevent performance deterioration. Interestingly, although AlexNet saw a substantial increase in accuracy when trained with StyleGANv3-generated data (+31.75%), it performed worse under the Guided-Diffusion condition, implying a sensitivity to the synthetic data's distributional shifts. Overall, the findings suggest that the success of data augmentation strategies is highly model-dependent, with DenseNet and MobileNetV2 demonstrating superior compatibility with both generative approaches. These results emphasize the importance of tailoring augmentation methods to model characteristics in order to optimize classification outcomes.

It can be observed that DenseNet, when trained with StyleGANv3-augmented data, achieved the highest accuracy of 91.38%, followed closely by MobileNetV2 with 88.15% accuracy under the Guided-Diffusion setting. Both results reflect considerable improvements over their respective vanilla models, which scored 86.10% and 85.61%. These findings suggest that certain architectures-particularly DenseNet and MobileNetV2-can benefit substantially

Table 8 Ablation study on synthetic data augmentation using StyleGAN3 and Diffusion models

Training set	Gazi Brains 2025			BRaTS21		
	Precision	Recall	Accuracy	Precision	Recall	Accuracy
StyleGANv3 Augmentation						
vanilla	84.32	88.20	86.10	75.71	77.27	78.22
vanilla + 200 synthetic	90.32 ± 2.88	90.54 ± 3.43	91.21 ± 2.70	81.88 ± 3.22	82.19 ± 4.15	81.39 ± 4.78
vanilla + 300 synthetic	91.41 ± 2.23	91.63 ± 3.10	92.29 ± 1.98	82.74 ± 3.22	83.47 ± 4.15	82.33 ± 4.78
vanilla + 480 synthetic	93.72 ± 3.52	94.04 ± 2.41	94.81 ± 1.35	85.38 ± 3.22	85.79 ± 4.15	84.79 ± 4.78
Diffusion Augmentation						
vanilla	84.32	88.20	86.10	76.04	76.85	77.41
vanilla + 200 synthetic	89.78 ± 3.12	91.02 ± 3.67	90.43 ± 2.49	78.88 ± 5.78	82.19 ± 4.69	78.45 ± 6.14
vanilla + 300 synthetic	92.58 ± 2.61	90.97 ± 3.48	93.41 ± 2.27	79.74 ± 4.53	78.34 ± 5.98	79.88 ± 6.58
vanilla + 480 synthetic	95.31 ± 3.89	93.12 ± 2.03	95.02 ± 1.94	81.44 ± 5.16	80.79 ± 4.41	81.88 ± 7.42

Best-performing architectures (DenseNet for Gazi Brains 2025, ResNeXt50 for BRaTS21) were fine-tuned with 200, 300, and 480 synthetic images. Results show mean ± std. dev. across 5 independent runs. Augmentation with 480 synthetic images achieves the highest accuracy: 94.81% (Gazi Brains 2025, StyleGAN3) and 84.79% (BRaTS21, StyleGAN3), demonstrating substantial improvements over vanilla baselines

Table 9 Hyperparameter search space and optimal configuration selected based on testing accuracy

Hyperparameter	Search Space
Batch Size	{16, 32}
Learning Rate	{0.0001, 0.0005, 0.001}
Optimizer	{Adam, RMSprop}
Dropout Rate	{0.0, 0.5}
Weight Decay	$\{0.0, 1\times10^{-5}\}$
Selected Parameters	**bs=16, lr=0.0001, Adam, do=0.5, wd=0.0**

Bold texts indicate the best values

Table 10 Multiclass classification (7-class) results for the Gazi Brains 2025 dataset

Model	Loss	Acc.	Prec.	Recall	F1-score
AlexNet	1.6772	0.3202	0.1025	0.3202	0.1553
MobileNetV2	1.1393	0.6601	0.6692	0.6601	0.6593
EfficientNetV2S	**1.0497**	**0.7233**	**0.7183**	**0.7233**	**0.7155**
ResNeXt50	1.0404	0.6403	0.6251	0.6403	0.6236
DenseNet	**0.8195**	**0.7075**	**0.7212**	**0.7145**	**0.7031**
Vision Transformer	1.3581	0.5460	0.5373	0.5460	0.4878
DINOv2-S	1.3867	0.4822	0.4767	0.4822	0.4244
DINOv2-L	1.2066	0.5697	0.5558	0.5697	0.5420

Bold texts indicate the best values

Model performance in distinguishing seven distinct neurological conditions, measured via loss, accuracy, precision, recall, and F1-score

from synthetic data generated via GANs and diffusion models. Significant improvements were also recorded in models such as EfficientNetV2S (accuracy increased to 91.81% with StyleGANv3) and ResNeXt50_32x4d (accuracy increased to 89.66% with StyleGANv3). However, not all models benefitted from augmentation. For example, VGG19 experienced a drop in performance with diffusion-based augmentation, indicating that augmentation efficacy may be model-dependent. Figures 6, 7, 8 also show performance metrics of vanilla, StyleGANv3-based, and diffusion-based models on the Gazi Brains 2025 dataset.

Based on the ablation study presented in Table 7, the integration of synthetic data generated by BetaVAE yields modest improvements in classification performance. The baseline vanilla models, trained without synthetic data, are compared against models fine-tuned with additional sets of 200, 300, and 480 synthetic images. For the Gazi Brains 2025 dataset, the inclusion of 200 synthetic images resulted in the highest mean accuracy (86.61%) and precision (89.27%), while the largest set of 480 synthetic images also provided a noticeable accuracy improvement over the baseline. For the BRaTS21 dataset, the greatest mean performance gains across precision, recall, and accuracy were similarly observed with the addition of 200 synthetic images. The results indicate that performance does not scale consistently with the quantity of synthetic data, as the largest synthetic set (480) did not yield the highest metrics. Furthermore, the performance achieved with BetaVAE-augmented data was lower than that obtained in comparable runs using synthetic data from StyleGAN3 and Guided Diffusion models.

Ablation study on the effect of synthetic data augmentation for brain MRI classification is given in Table 8. We evaluate the impact of augmenting the vanilla training set with varying amounts of synthetic images generated by StyleGAN3 and Diffusion models. The best-performing architectures (DenseNet for Gazi Brains 2025, ResNeXt50 for BRaTS21) were selected after hyperparameter tuning and used for fine-tuning with synthetic data. The hyperparameter search space is provided in Table 9. Results are reported as mean ± standard deviation across 5 independent runs with fixed random seeds. Progressive augmentation with 200, 300, and 480 synthetic images demonstrates consistent improvements in precision, recall, and accuracy for both datasets. The most substantial gains are observed with 480 synthetic images, achieving 94.81% (±1.35%) accuracy on Gazi Brains 2025 with StyleGAN3 augmentation and 95.02% (±1.94%) with Diffusion augmentation,

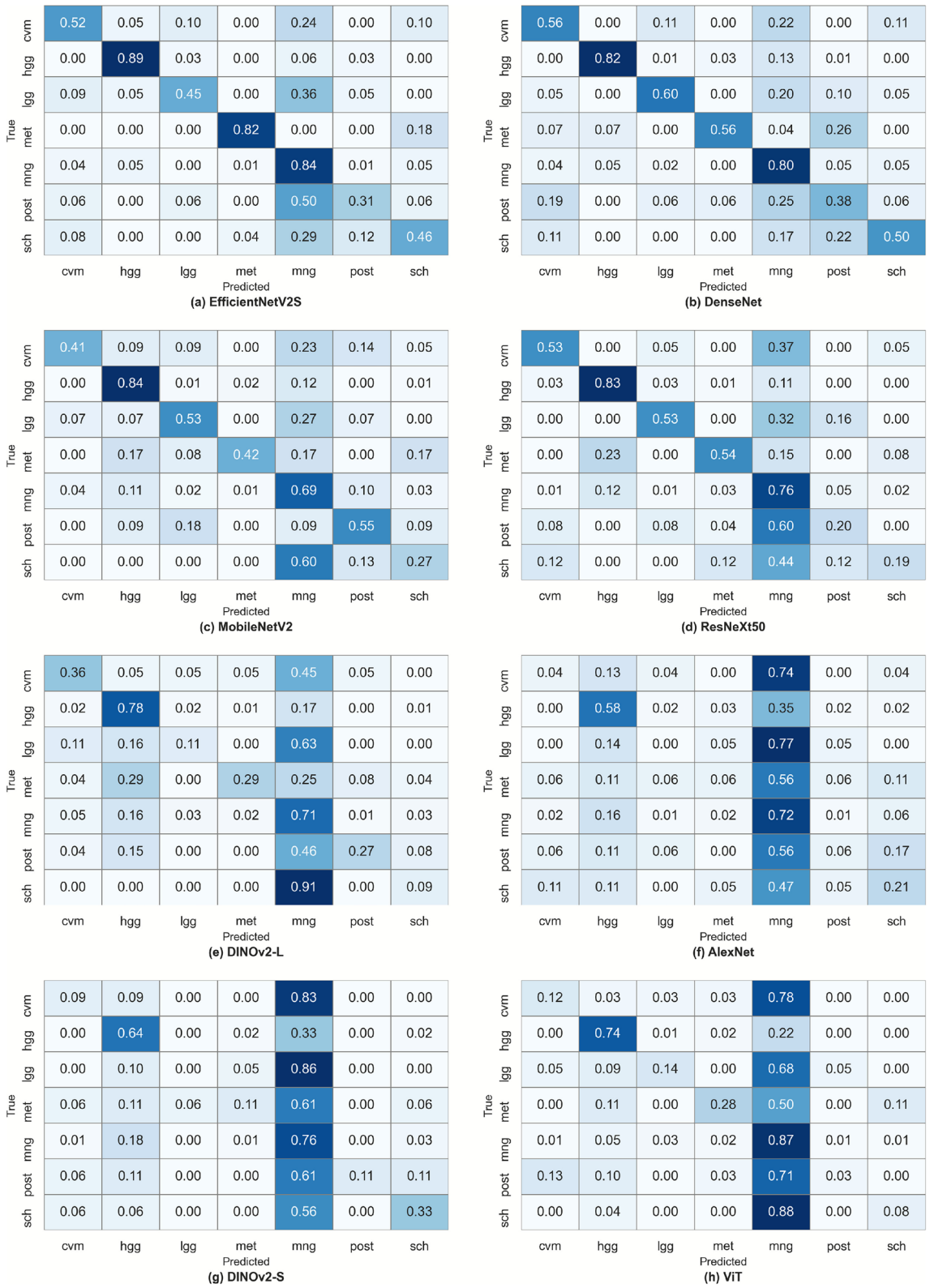


Fig. 9 Confusion matrices for multiclass classification (7-class) models on the Gazi Brains 2025 dataset

Table 11 Comparative analysis of synthetic MRI quality metrics (FID, EQ-T, EQ-R) and computational efficiency between Gazi Brains 2025 and BraTS 2021 neuroimaging datasets

Dataset	FID (↓)		EQ-T (↑)		EQ-R (↑)		Duration
	Val.	Dur.	Val.	Dur.	Val.	Dur.	(d:h:m)
Gazi Brains 2025	26.2	13:43	71.2	50:37	18.2	50:58	6:06:07
BraTS 2021	34.6	13:44	72.8	50:46	14.3	51:03	6:01:59

Table 12 Guided-Diffusion synthetic image quality metrics across datasets

Dataset	IS	FID	sFID	Train	Sample	Eval.
				(d:h:m)	(d:h:m)	(m:s)
Gazi Brains 2025	1.59	8.66	93.89	4:19:30	3:23:21	3:39
BraTS 2021	3.18	45.60	66.23	4:18:01	0:23:48	5:04

Comprehensive quality analysis (IS, FID, sFID) and computational costs for diffusion-generated images

representing 8.71% and 8.92% absolute improvements over the vanilla baseline (86.10%), respectively. For BRaTS21, improvements of 6.57% and 4.47% are observed with StyleGAN3 and Diffusion augmentation, respectively, demonstrating the effectiveness of generative data augmentation across different datasets and model architectures.

4.6 Multi-class classification results for brain tumors

A comprehensive 7-class classification task is carried out to evaluate the performance of various deep learning architectures in accurately identifying distinct neurological pathologies within the Gazi Brains 2025 dataset at this stage of the research. Both convolutional and transformer-based architectures are systematically examined to assess their capability to extract meaningful features and effectively distinguish between multiple categories of brain abnormalities. Model performance is quantitatively evaluated using a set of standard metrics, including categorical cross-entropy loss, overall accuracy, class-wise precision, recall, and F1-score. A detailed comparison of the results achieved by each architecture is presented in Table 10, offering insights into their respective strengths and diagnostic effectiveness. Figure 9 shows the confusion matrices for multiclass classification (7-class) models on the Gazi Brains 2025 dataset.

Table 10 provides a comprehensive evaluation of multiple deep learning architectures applied to a seven-class classification task using the Gazi Brains 2025 dataset. Among the models assessed, EfficientNetV2S demonstrated superior overall performance, yielding the lowest classification loss (1.0497) and the highest accuracy (72.33%). Additionally, it attained the highest precision (0.7183) and recall (0.7233) values, culminating in an F1-score of 0.7155. These results underscore EfficientNetV2S's strong capability to generalize across diverse neurological categories, offering a balanced trade-off between sensitivity and specificity.

DenseNet closely followed in performance, achieving an accuracy of 70.75%, with precision, recall, and F1-score values of 0.7212, 0.7145, and 0.7031, respectively. Its relatively low loss (0.8195) further highlights its reliability in handling complex multi-class differentiation tasks within neuroimaging data. MobileNetV2 produced moderate classification outcomes, attaining an accuracy of 66.01% and an F1-score of 0.6593. Although not as robust as EfficientNetV2S or DenseNet, MobileNetV2 maintained a relatively balanced performance profile, indicating its potential utility in computationally constrained environments where efficiency is prioritized.

Intermediate results were observed for ResNeXt50 and DINOv2-L, which recorded accuracies of 64.03% and 56.97%, respectively. While ResNeXt50 exhibited consistent precision and F1-score values (0.6251 and 0.6236), DINOv2-L's performance was somewhat limited, particularly in terms of F1-score (0.5420), suggesting room for improvement in learning nuanced inter-class representations.

Conversely, models such as Vision Transformer, DINOv2-S, and especially AlexNet yielded suboptimal performance. Vision Transformer achieved an accuracy of 54.60% and an F1-score of 0.4878, while DINOv2-S lagged further behind (accuracy: 48.22%, F1-score: 0.4244). AlexNet performed the worst, with only 32.02% accuracy and an F1-score of 0.1553, emphasizing its limitations in high-complexity, multi-class medical image classification settings.

The relatively modest performance across several models can be attributed to the class imbalance present within the Gazi Brains 2025 dataset, which contains a disproportionate number of slices from dominant classes such as High-Grade Glioma and normal cases. This imbalance likely hindered the models' ability to adequately learn minority class representations, thereby negatively impacting precision and F1-scores.

In conclusion, while architectures such as EfficientNetV2S and DenseNet emerge as promising candidates for multi-class classification in brain MRI analysis, the diminished performance of transformer-based and shallow CNN architectures points to the need for further architectural fine-tuning and the application of advanced class-balancing strategies during training.

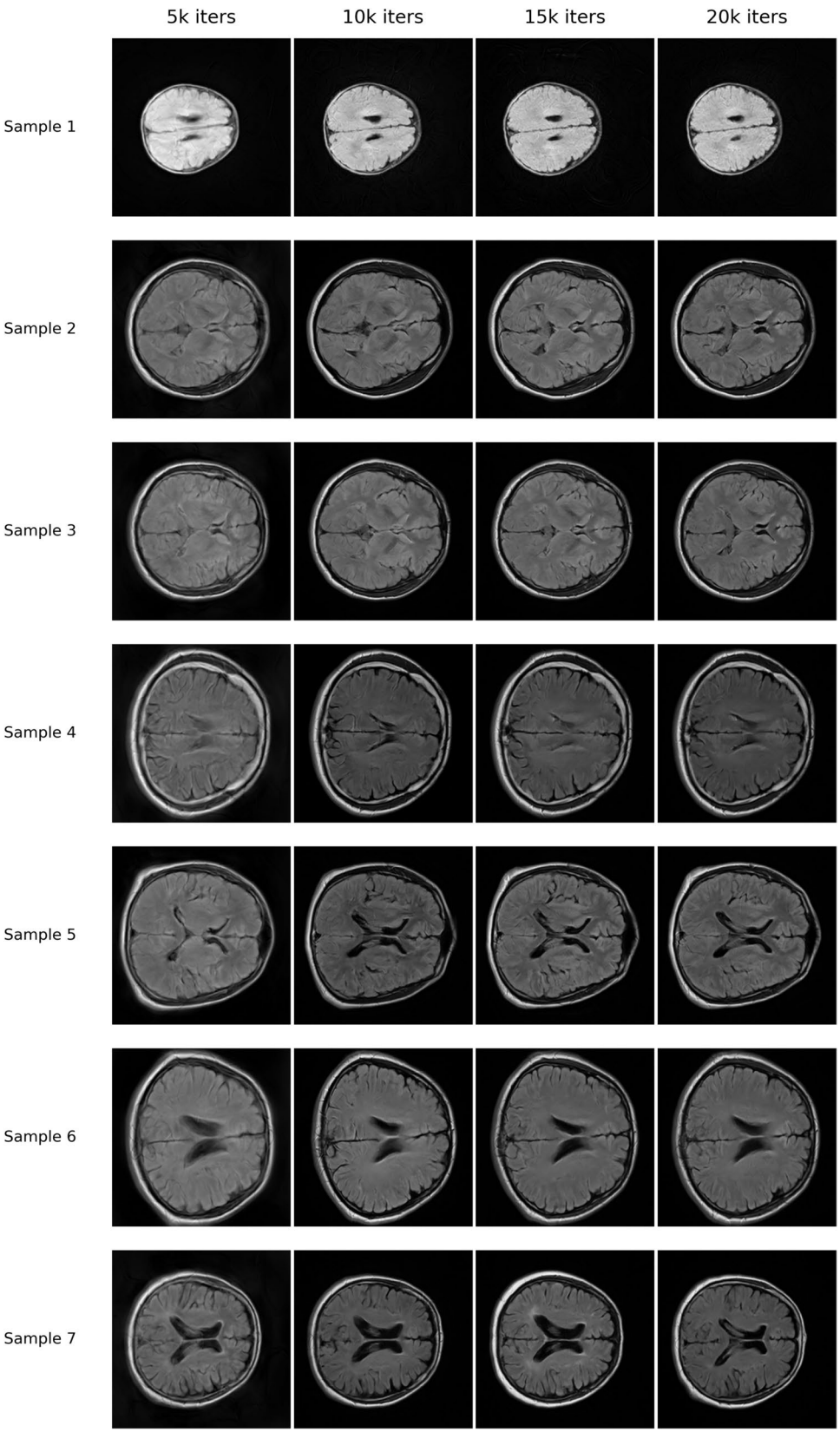


Fig. 10 Progressive quality improvement of StyleGAN3-generated brain MRI images across training iterations for Gazi Brains 2025 Dataset

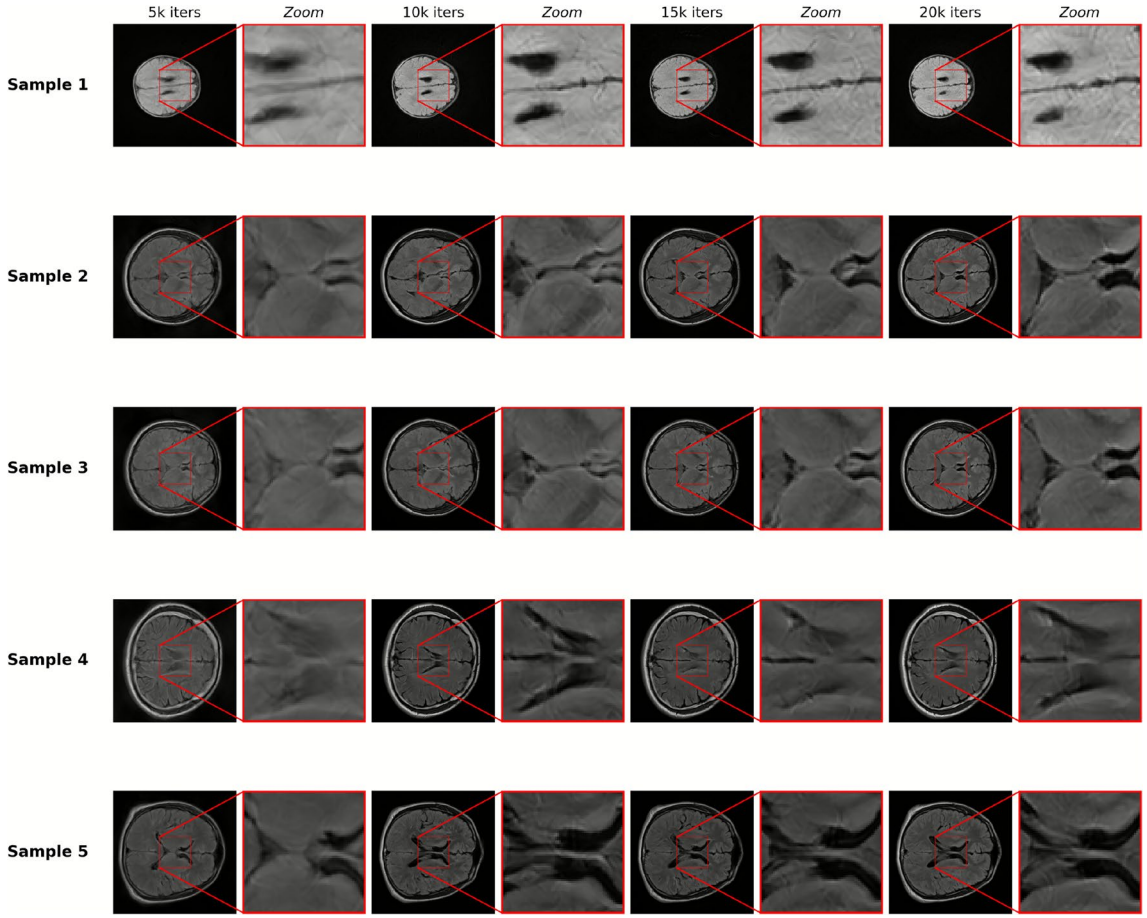


Fig. 11 Progressive evolution of StyleGAN3-generated brain MRI samples from Gazi Brains dataset across training iterations (5k–20k). Each row shows the same sample at different training stages, with magnified regions (red boxes) revealing increasing detail fidelity and artifact reduction as training progresses

4.7 Generative model evaluation

In this section, the performance of the StyleGANv3-T configuration is evaluated using the Gazi Brains 2025 and BraTS 2021 datasets to assess the realism, diversity, and representational consistency of the generated synthetic MRI slices. Three distinct quantitative metrics are used: FID, EQ-T, and EQ-R, in addition to reporting the total training duration for each dataset. The results are given in Table 11.

Lower values of FID indicate higher image quality and better alignment with the real data distribution. The StyleGANv3-T model achieved a FID of 26.21 when trained on the Gazi Brains 2025, which is notably lower than the FID score of 34.57 obtained from the BRaTS21 dataset. This result indicates that the images synthesized from Gazi Brains 2025 exhibit greater visual realism and more faithfully replicate the target anatomical features present in the real dataset.

EQ-T and EQ-R, on the other hand, serve to evaluate the structural and representational consistency of the generated samples. EQ-T measures topological faithfulness in low-dimensional embeddings, reflecting how well the generated samples preserve the local relationships observed in the original dataset. EQ-R quantifies how representative the generated samples are in terms of occupying the real data's embedding space. BraTS 2021 exhibited slightly higher EQ-T (72.79) compared to Gazi Brains 2025 (71.21), suggesting that the synthetic data derived from BraTS 2021 retained more consistent local structure. However, Gazi Brains 2025 achieved a higher EQ-R value (18.16) relative to BraTS 2021 (14.34), indicating that its synthetic samples more thoroughly and diversely represented the feature space of the original dataset. These results highlight a nuanced trade-off between image realism and topological consistency.

In summary, StyleGANv3-T yielded visually higher-quality synthetic MRI images for the Gazi Brains 2025 dataset, while BraTS 2021 demonstrated marginally stronger topological structure retention. These outcomes reinforce the adaptive generative performance of StyleGANv3 in relation to dataset complexity and variability.

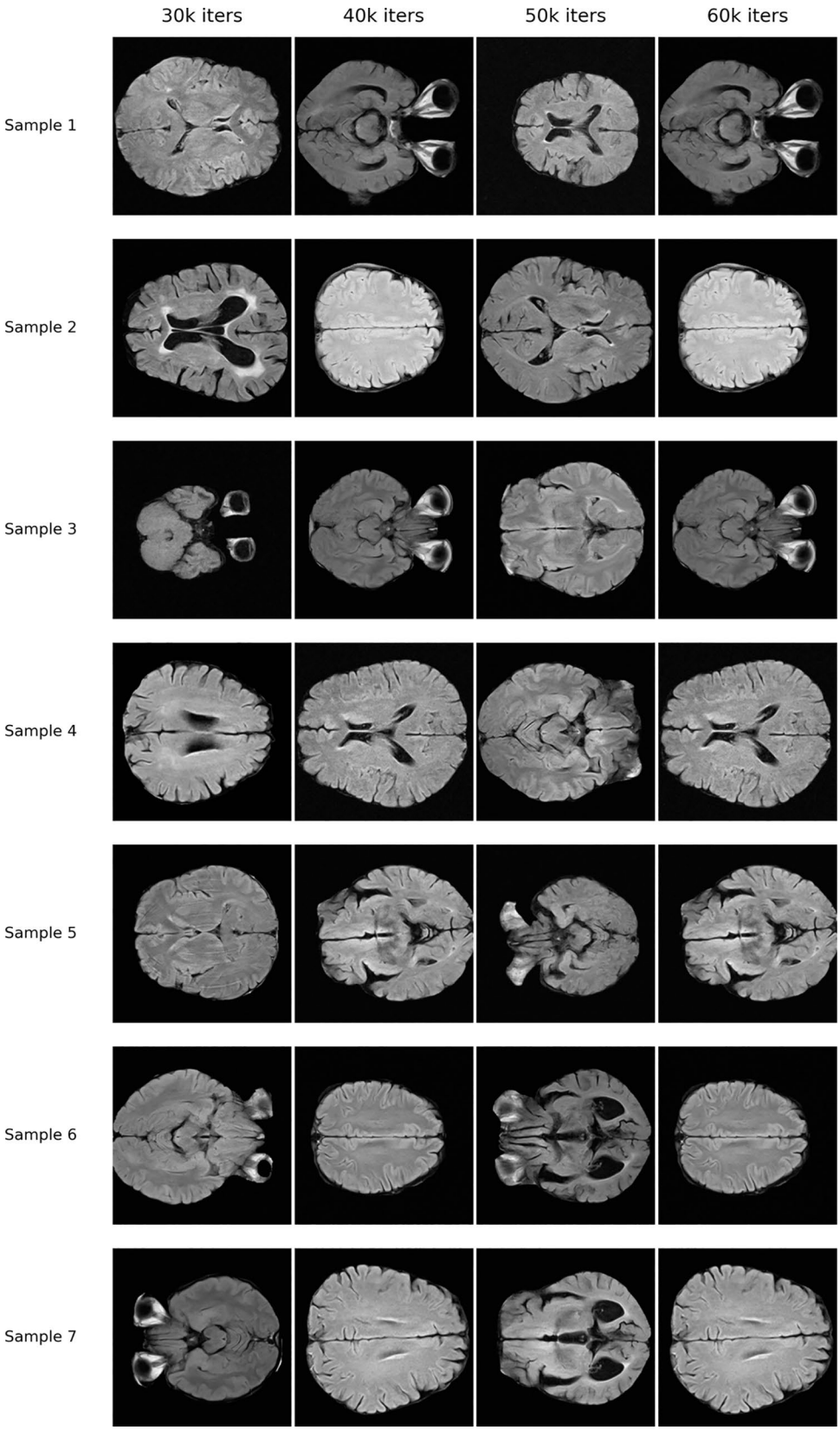


Fig. 12 Progressive quality improvement of diffusion-generated brain MRI images across training iterations for Gazi Brains 2025 Dataset

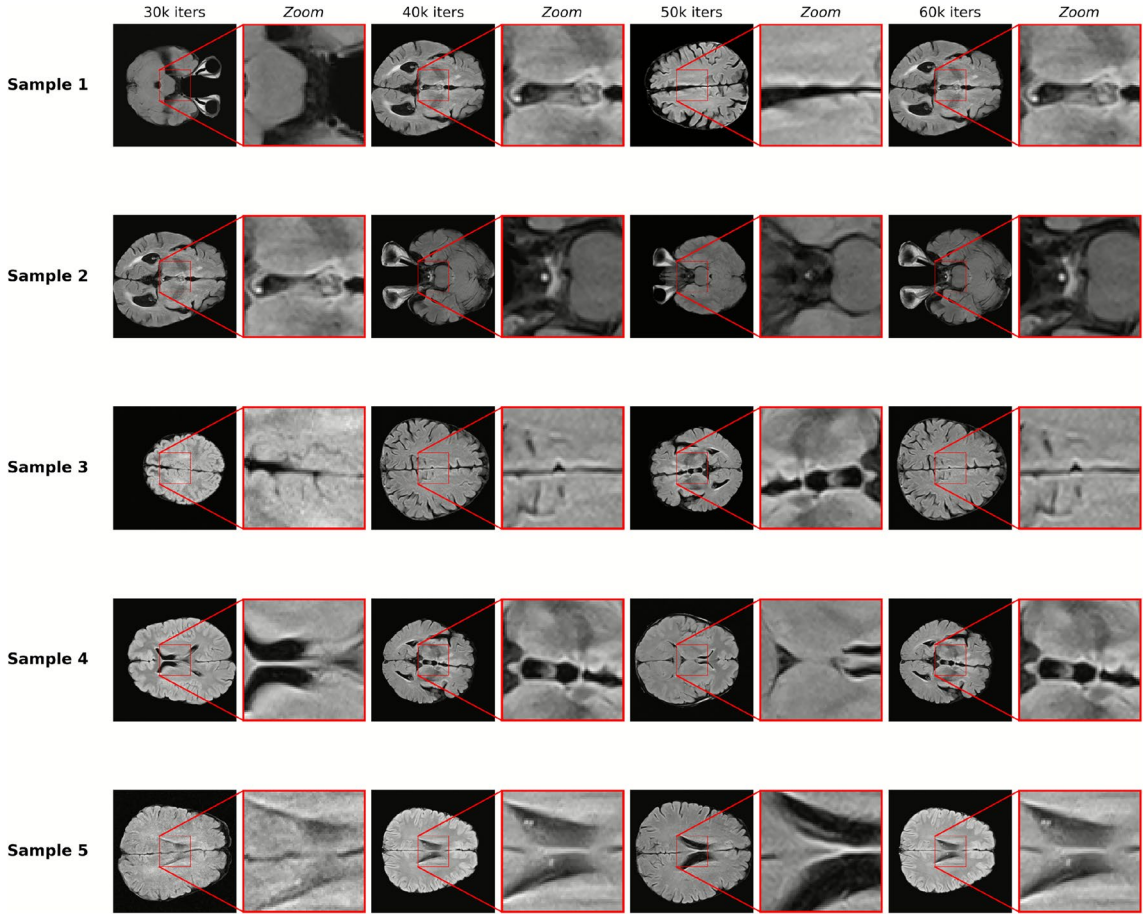


Fig. 13 Progressive evolution of diffusion-generated brain MRI samples from Gazi Brains dataset across training iterations (40k–60k). Each row shows the same sample at different training stages, with magnified regions (red boxes) revealing increasing detail fidelity and artifact reduction as training progresses

The performance of the Guided Diffusion model has been evaluated on the both Gazi Brains 2025 and BraTS 2021 datasets, with the results summarized in Table 12. Several well-established quantitative metrics were employed to assess image quality. The Fréchet Inception Distance (FID) measures the similarity between the real and generated data distributions, where lower scores indicate closer alignment and, therefore, more realistic outputs. The sliced FID (sFID) is a variant of FID that considers lower-dimensional projections to provide robustness against overfitting or mode collapse. Again, lower values are desirable. The Inception Score (IS) evaluates the generated images based on their classifiability and diversity, with higher scores reflecting better performance.

In addition, precision quantifies the proportion of generated images that fall within the distribution of real images (realism), while recall measures the extent to which the model can reproduce the diversity of the real dataset (coverage).

The results indicate that the Guided Diffusion model trained on the Gazi Brains 2025 dataset achieved markedly superior realism and diversity compared to the BraTS 2021 counterpart. A substantially lower FID of 8.66 and a higher recall of 0.9518 suggest that the model was highly effective in capturing the underlying distribution of the Gazi Brains 2025 MRI images. In contrast, the BraTS 2021 model produced images with a significantly higher FID of 45.60, reflecting reduced visual fidelity, and a lower recall of 0.3941, indicating weaker representation of the original data's variability.

Interestingly, the Inception Score (IS) for BraTS 2021 (3.18) was higher than that of Gazi Brains 2025 (1.59), which may reflect greater intra-class diversity in the generated BraTS 2021 images; however, this does not necessarily correlate with visual or clinical quality, especially in the context of medical imaging where realism and anatomical correctness are paramount.

Overall, the experimental findings underscore the effectiveness of the Guided Diffusion model in generating high-resolution, diverse, and realistic MRI images when trained on the Gazi Brains 2025 dataset. This outcome not only reflects the model's robust generative capabilities but also

highlights the importance of high-quality datasets in facilitating effective learning. The superior precision and recall metrics further indicate that the dataset enabled the model to better capture underlying data distributions, leading to improved generalization. These results collectively suggest that diffusion-based models hold significant promise for medical image synthesis, with potential applications in data augmentation, diagnostic decision support, and broader clinical practice.

The Fig. 10 displays synthetic brain MRI images generated by StyleGAN3 at four training checkpoints: 5,000 (5k), 10,000 (10k), 15,000 (15k), and 20,000 (20k) iterations. Images within each row track the evolution of the same underlying structure across iterations, selected using mean squared error (MSE)–based similarity matching to ensure that consistent anatomical features are compared. Early iterations (5k) show coarse, low-quality outputs with visible artifacts, while later iterations (20k) produce increasingly realistic brain anatomy with improved tissue contrast, structural coherence, and reduced noise. This progression demonstrates the generator's learned capacity to synthesize anatomically plausible brain MRI scans. All images are 256×256 pixels displayed in grayscale. Additionally, zoomed-in versions of the 5–20k examples are provided in a separate figure, shown in Figure 11, to highlight how fine-grained structural details improve over the course of training.

Similarly, synthetic MRI samples produced by the Guided-Diffusion model are presented in Fig. 12. These samples illustrate the progressive refinement achieved throughout the diffusion denoising steps, where earlier stages contain noisy intermediate structures and later stages converge to highly realistic brain anatomy with sharper boundaries and clearer tissue differentiation. Consistent anatomical regions are compared across diffusion steps using the same MSE-based similarity matching strategy applied to the StyleGAN3 outputs. Zoomed views of the Guided-Diffusion samples, provided in Fig. 13, further highlight how subtle structural features-such as cortical textures, ventricle definition, and tissue transitions-become increasingly coherent as the model denoises toward the final reconstruction.

5 Discussion and conclusion

A new dataset, Gazi Brains 2025, consisting of MRI scans from 500 patients, is introduced in this study. A significant portion of these scans shows brain anomalies, which are annotated and preprocessed to support both binary and multiclass classification tasks. Several deep learning and transformer-based models are developed to classify the images as either tumor-positive or normal, as well as to distinguish among seven neurological conditions. Synthetic data augmentation is applied using StyleGANv3 and Guided Diffusion to examine whether such techniques can improve classifier performance.

The experimental studies demonstrated that synthetic augmentation improved classification accuracy compared to baseline models trained only on the original dataset. As shown in Tables 5 and 6, both StyleGANv3-based and Guided Diffusion-based augmentation led to increased accuracy, with models such as DenseNet and MobileNetV2 achieving close to 91% accuracy in the binary classification task. However, the improvement in accuracy comes with a high computational cost. For example, StyleGANv3 required over six days of training, while Guided Diffusion took nearly nine days in total, including training and sampling. This extended processing time is a drawback for practical applications where quicker turnaround is needed.

In addition to the main performance evaluation, an extensive hyperparameter search was conducted to ensure fair comparison across models, as summarized in Table 9. After identifying the optimal configurations, the best-performing architectures from Tables 6 and 5 were selected for a more fine-grained ablation study. The ablation results presented in Table 8 demonstrate the effect of incremental synthetic augmentation (+200, +300, +480 samples) on classification performance for both StyleGANv3 and Guided Diffusion. This analysis provides further evidence that performance gains are not incidental; rather, they systematically increase with additional high-quality synthetic data, reinforcing the role of generative augmentation as a core contributing factor to the improved results.

A comparison between the two augmentation techniques shows that each has its own trade-offs. The StyleGANv3 approach produces synthetic images with a low Fréchet Inception Distance (FID), indicating that the generated images are close to real images in the feature space. However, the architecture of StyleGANv3 limits the number of samples it can generate. In contrast, the diffusion-based method can produce a larger volume of synthetic data, which is useful for covering the variability in the dataset. Despite this advantage, the Guided Diffusion model requires significantly more time for training and sampling.

These results must also be compared with our previous work on the smaller Gazi Brains 2020 dataset, which included MRI scans from 100 patients. In that study, augmentation using StyleGANv2 resulted in classification accuracies as high as 98%. In the current study, the highest accuracy obtained was around 91%. This difference may be due to the broader range of anomalies present in the Gazi Brains 2025 dataset, which leads to an uneven distribution of cases. The imbalance in class representation likely contributed to the lower performance observed.

In summary, synthetic augmentation using both StyleGANv3 and Guided Diffusion has been shown to improve classifier accuracy on the Gazi Brains 2025 dataset, but at the expense of considerable training and sampling times. The diffusion-based approach can generate larger synthetic datasets compared to StyleGANv3, though it requires more computational resources. The lower accuracy in this study relative to the previous work on the Gazi Brains 2020 dataset is likely related to the increased heterogeneity and imbalance of the current dataset. Future work should focus on reducing computational time and addressing class imbalance to further enhance classification performance in this setting.

To clarify the distinct contributions of this study and how it advances existing methods, the key elements of the proposed approach are summarized below:

- A new MRI dataset, Gazi Brains 2025, containing 500 patients, is introduced.
- Annotated scans support both binary and multiclass classification tasks.
- Deep learning and transformer-based classifiers are developed for seven neurological conditions.
- Synthetic augmentation using StyleGANv3 and Guided Diffusion is integrated into the classification pipeline.
- The primary motivation for generative augmentation is to address limited dataset size and severe class imbalance.
- Data scarcity is a major issue in medical imaging, and synthetic augmentation mitigates this limitation.
- Synthetic augmentation improves classification accuracy compared to models trained only on original data.
- Accuracy gains are demonstrated across multiple architectures, including DenseNet and MobileNetV2.
- The study evaluates the trade-off between accuracy improvements and the high computational cost of generative models.
- A comparative analysis of two augmentation techniques is conducted within a unified experimental framework.
- The work introduces an end-to-end workflow combining dataset creation, preprocessing, augmentation, and multi-model evaluation.
- The study shows how synthetic augmentation contributes to enhanced performance in medical image classification tasks.

5.1 Practical implications

The findings of this study have significant practical implications for the development of computer-aided diagnostic tools in medical imaging, particularly for brain anomaly detection using MRI. By introducing the Gazi Brains 2025 dataset, this research provides a valuable resource for training and benchmarking machine learning models in real-world clinical scenarios. The inclusion of the both binary and multiclass classification tasks allows researchers and developers to target a broad spectrum of diagnostic challenges, ranging from simple tumor detection to differentiating between multiple neurological disorders.

Furthermore, the comparison between StyleGANv3 and diffusion-based methods provides critical insights for practitioners choosing an augmentation strategy. While StyleGANv3 offers high-fidelity samples with lower FID scores, its limited scalability may not meet the demands of highly imbalanced or heterogeneous datasets. Conversely, Guided Diffusion is more suitable for generating large-scale augmentations, especially when aiming to balance underrepresented classes. These trade-offs inform practical decisions about model deployment and data augmentation in clinical workflows.

5.2 Limitations and future works

Despite the promising results, several limitations should be acknowledged. The study remains dependent on the characteristics of the Gazi Brains 2025 and BraTS21 datasets, and performance may vary when applied to datasets with different imaging protocols or population distributions. The generative models used for augmentation-StyleGANv3 and Guided Diffusion-are computationally demanding, requiring substantial GPU resources and long training times, which may limit their applicability in resource-constrained environments. Additionally, both the augmentation and classification stages exhibit sensitivity to hyperparameter configurations, requiring careful tuning to achieve stable performance. These factors collectively indicate that further work is needed to improve generalizability, reduce computational cost, and enhance the robustness of the pipeline.

Another limitation lies in the class imbalance and heterogeneity of the Gazi Brains 2025 dataset. Although the dataset provides a more realistic and diverse representation of neurological anomalies compared to previous versions, the uneven distribution of classes negatively impacted multiclass classification performance. Additionally, while the study demonstrates improvements in classification accuracy due to synthetic augmentation, it does not explore the clinical interpretability or explainability of the models. In real-world healthcare applications, it is crucial that AI systems not only perform well but also provide interpretable outputs that clinicians can trust. This aspect was not addressed in the current work and remains an important direction for future research. Finally, the experiments were conducted on a single dataset collected from one institution, which may limit the generalizability of the findings. External validation on datasets from different hospitals or imaging devices was not

performed, which is essential to assess the robustness of the models in diverse clinical environments.

We seek to address the computational inefficiencies inherent in current generative models by investigating more efficient and lightweight augmentation techniques. Future efforts will focus on the exploration of advanced strategies, such as diffusion model distillation and hybrid GAN-diffusion frameworks, with the goal of enhancing both the computational efficiency and practical applicability of generative approaches in data augmentation pipelines. We also plan to implement class-balancing strategies, including targeted synthetic sampling and loss-based reweighting, to mitigate the impact of data imbalance observed in the Gazi Brains 2025 dataset. In addition, integrating explainable AI (XAI) methods will be a priority to enhance the interpretability and clinical trustworthiness of the models. To improve generalizability, we intend to validate our models on external datasets from multiple institutions and investigate domain adaptation techniques to ensure consistent performance across varied imaging sources and patient populations.

Acknowledgements This paper was prepared for a part of the Turkish Brain Project (No. 2021-001) and supported by the Turkish Presidency Digital Transformation Office (TR-DTO). The authors would like to thank TR-DTO for their financial support; Gazi University Medical School Hospital for providing facilities; and the colleagues from TR-DTO who provided preliminary infrastructure and support for this research project.

Author contributions E.C., A.O.B, A.S., M.C.S., Z.D., and K.N. are primarily responsible for data collection, acquisition, clinical annotation, and verification of the dataset. Their medical expertise ensured the reliability, diagnostic relevance, and clinical accuracy of the collected imaging data. M.M.E.Y., S.E., B.A., Y.A., and Ş.S. focused on the experimental setup, deep learning modeling, training, testing, and validation procedures, as well as computational aspects of the research. They were also responsible for data preprocessing, privacy preserving, and conducting the literature review. Additionally, they led the design and implementation of the end-to-end system pipeline, including software development, model training and validation, and performance evaluation. All contributors further managed manuscript preparation and refinement, and coordinated the peer-review process, ensuring technical coherence and methodological rigor throughout the publication workflow.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). Turkish Brain Project was finally supported by the Turkish Presidency Digital Transformation Office (TR-DTO).

Data Availability Data is available upon request due to privacy/ethical restrictions.

Code availability Code is available at https://github.com/meyurtsever/gazi-brains-mri-classification/

Declarations

Competing interests The authors declare no competing interests.

Ethical approval Ethical approval was obtained from the ethics committee at Gazi University.

Consent to participate The authors confirm that informed consent has been obtained from the involved patients or if appropriate from the parent, guardian, or power of attorney of the involved patients; and, they have given approval for this information to be published in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Feigin, V.L., Nichols, E., Alam, T., Bannick, M.S., Beghi, E., Blake, N., et al.: Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol. **18**(5), 459–480 (2019)
2. Li, X., Ma, Y., Fan, Z., Khandia, R.: Brain Tumors: Advancements in Diagnostics and Innovative Therapies. CRC Press (2024)
3. Mathivanan, S.K., Sonaimuthu, S., Murugesan, S., Rajadurai, H., Shivahare, B.D., Shah, M.A.: Employing deep learning and transfer learning for accurate brain tumor detection. Sci. Rep. **14**(1), 7232 (2024)
4. Srinivasan, S., Francis, D., Mathivanan, S.K., Rajadurai, H., Shivahare, B.D., Shah, M.A.: A hybrid deep cnn model for brain tumor image multi-classification. BMC Med. Imaging **24**(1), 21 (2024)
5. Yurtsever, M.E., Atay, Y., Arslan, B., Sagiroglu, S.: Development of brain tumor radiogenomic classification using gan-based augmentation of mri slices in the newly released gazi brains dataset. BMC Med. Inform. Decis. Mak. **24**(1), 285 (2024)
6. Abd-Ellah, M.K., Awad, A.I., Khalaf, A.A., Hamed, H.F.: A review on brain tumor diagnosis from mri images: practical implications, key achievements, and lessons learned. Magn. Reson. Imaging **61**, 300–318 (2019)
7. Irmak, E.: Multi-classification of brain tumor mri images using deep convolutional neural network with fully optimized framework. Iran. J. Sci. Technol. Trans. Electr. Eng. **45**(3), 1015–1036 (2021)
8. Mehnatkesh, H., Jalali, S.M.J., Khosravi, A., Nahavandi, S.: An intelligent driven deep residual learning framework for brain tumor classification using mri images. Expert Systems with Applications **213**, 119087 (2023)
9. Asadi, F., Angsuwatanakul, T., O'Reilly, J.A.: Evaluating synthetic neuroimaging data augmentation for automatic brain tumour segmentation with a deep fully-convolutional network. BRO Neurosci. Rep. **16**, 57–66 (2024)
10. Pani, K., Chawla, I.: Synthetic mri in action: a novel framework in data augmentation strategies for robust multi-modal brain tumor segmentation. Comput. Biol. Med. **183**, 109273 (2024)

11. Mahnoor, M., Rainio, O., Klén, R.: Exploring generative adversarial network-based augmentation of magnetic resonance brain tumor images. Applied Sciences **14**(24), 11822 (2024)
12. Gab Allah, A.M., Sarhan, A.M., Elshennawy, N.M.: Classification of brain mri tumor images based on deep learning pggan augmentation. Diagnostics **11**(12), 2343 (2021)
13. Yapici, M., Karakis, R., Gurkahraman, K.: Improving brain tumor classification with deep learning using synthetic data. Computers, Materials and Continua **74**(3), (2023)
14. Alrashedy, H.H.N., Almansour, A.F., Ibrahim, D.M., Hammoudeh, M.A.A.: Braingan: brain mri image generation and classification framework using gan architectures and cnn models. Sensors **22**(11), 4297 (2022)
15. Nag, A., Mondal, H., Hassan, M.M., Al-Shehari, T., Kadrie, M., Al-Razgan, M., Bairagi, A.K.: Tumorganet: A transfer learning and generative adversarial network-based data augmentation model for brain tumor classification. IEEE Access (2024). https://doi.org/10.1109/access.2024.3429633
16. Ghassemi, N., Shoeibi, A., Rouhani, M.: Deep neural network with generative adversarial networks pre-training for brain tumor classification based on mr images. Biomed. Signal Process. Control **57**, 101678 (2020)
17. Trabucco, B., Doherty, K., Gurinas, M., Salakhutdinov, R.: Effective data augmentation with diffusion models. (2023). Preprint at arXiv:2302.07944
18. Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., Cattin, P.C.: Diffusion models for implicit image segmentation ensembles. International conference on medical imaging with deep learning 1336–1348 (2022)
19. Gao, T., Hu, W., Chen, M., Chen, L., Jiang, H.: Msdmat-bts: multi-scale diffusion model and attention mechanism for brain tumor segmentation. Biomed. Signal Process. Control **104**, 107505 (2025)
20. Tao, T., Wang, Z., Zhang, H., Arvanitis, T.N., Zhang, L.: Diffkan-inpainting: Kan-based diffusion model for brain tumor inpainting. (2025). Preprint at arXiv:2502.16771
21. Dorjsembe, Z., Pao, H-. K., Odonchimed, S., Xiao, F.: Conditional diffusion models for semantic 3d brain mri synthesis. IEEE J. Biomed. Health Inform. (2024).
22. Fan, K., Cai, X., Niranjan, M.: Discrepancy-based diffusion models for lesion detection in brain mri. Comput. Biol. Med. **181**, 109079 (2024)
23. Zhao, H., Lou, H., Yao, L., Zhang, Y.: Diffusion transformer-augmented fmri functional connectivity for enhanced autism spectrum disorder diagnosis. J. Neural Eng (2025). https://doi.org/10.1088/1741-2552/adb07a
24. Liu, Z., Ma, C., She, W., Xie, M.: Innovative multi-class segmentation for brain tumor mri using noise diffusion probability models and enhancing tumor boundary recognition. Sci. Rep. **14**(1), 29576 (2024)
25. Kebaili, A., Lapuyade-Lahorgue, J., Vera, P., Ruan, S.: Multimodal mri synthesis with conditional latent diffusion models for data augmentation in tumor segmentation. Comput. Med. Imaging Graph **123**, 102532 (2025)
26. Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., Haworth, A.: A review of medical image data augmentation techniques for deep learning applications. J. Med. Imaging Radiat. Oncol. **65**(5), 545–563 (2021)
27. Goceri, E.: Medical image data augmentation: techniques, comparisons and interpretations. Artificial Intelligence Review **56**(11), 12561–12605 (2023)
28. Anaya-Isaza, A., Mera-Jiménez, L.: Data augmentation and transfer learning for brain tumor detection in magnetic resonance imaging. IEEE access **10**, 23217–23233 (2022)
29. Kollem, S., Reddy, K.R., Prasad, C.R., Chakraborty, A., Ajayan, J., Sreejith, S., Janapati, R.: Alexnet-ndtl: Classification of mri brain tumor images using modified alexnet with deep transfer learning and lipschitz-based data augmentation. Int. J. Imaging Syst. Technol. **33**(4), 1306–1322 (2023)
30. Mahmoud, A., Awad, N.A., Alsubaie, N., Ansarullah, S.I., Alqahtani, M.S., Abbas, M., Saber, A.: Advanced deep learning approaches for accurate brain tumor classification in medical imaging. Symmetry **15**(3), 571 (2023)
31. Rahman, T., Islam, M.S.: Mri brain tumor detection and classification using parallel deep convolutional neural networks. Measurement: Sensors **26**, 100694 (2023)
32. Basthikodi, M., Chaithrashree, M., Ahamed Shafeeq, B., Gurpur, A.P.: Enhancing multiclass brain tumor diagnosis using svm and innovative feature extraction techniques. Sci. Rep. **14**(1), 26023 (2024)
33. Bhagyalaxmi, K., Dwarakanath, B., Reddy, P.V.P.: Deep learning for multi-grade brain tumor detection and classification: a prospective survey. Multimedia Tools and Applications **83**(25), 65889–65911 (2024)
34. Yadav, A.C., Shah, K., Purohit, A., Kolekar, M.H.: Computer-aided diagnosis for multi-class classification of brain tumors using cnn features via transfer-learning. Multimedia Tools and Applications **84**(31), 38959–38982 (2025)
35. Zulfiqar, F., Bajwa, U.I., Mehmood, Y.: Multi-class classification of brain tumor types from mr images using efficientnets. Biomed. Signal Process. Control **84**, 104777 (2023)
36. Kibriya, H., Masood, M., Nawaz, M., Nazir, T.: Multiclass classification of brain tumors using a novel cnn architecture. Multimedia Tools and Applications **81**(21), 29847–29863 (2022)
37. Rastogi, D., Johri, P., Tiwari, V., Elngar, A.A.: Multi-class classification of brain tumour magnetic resonance images using multi-branch network with inception block and five-fold cross validation deep learning framework. Biomed. Signal Process. Control **88**, 105602 (2024)
38. Amin, M., Nahar, K.M., Gharaibeh, H., Nasayreh, A., Alsalmanc, N., Alomar, A., et al.: Diet transformer model with pca-ade integration for advanced multi-class brain tumor classification. Intell. Med. **11**, 100192 (2025)
39. Yadav, A.C., Kolekar, M.H., Sonawane, Y., Kadam, G., Tiwarekar, S., Kalbande, D.R.: Effunet++: A novel architecture for brain tumor segmentation using flair mri images. IEEE Access **12**, 152430–152443 (2024)
40. Sonawane, Y., Kolekar, M.H., Yadav, A.C., Kadam, G., Tiwarekar, S., Kalbande, D.R.: Dcrunet++: A depthwise convolutional residual unet++ model for brain tumor segmentation. In: Antonacopoulos, A., Chaudhuri, S., Chellappa, R., Liu, C.-L., Bhattacharya, S., Pal, U. (eds.) Pattern recognition, pp. 266–280. Springer Nature Switzerland, Cham (2025)
41. Yadav, A.C., Kolekar, M.H., Zope, M.K.: Modified recurrent residual attention u-net model for mri-based brain tumor segmentation. Biomedical Signal Processing and Control **102**, 107220 (2025)
42. Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage **31**(3), 1116–1128 (2006)
43. Karras, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., Lehtinen, J., Aila, T.: Alias-free generative adversarial networks. Adv. Neural. Inf. Process. Syst. **34**, 852–863 (2021)
44. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Adv. Neural. Inf. Process. Syst. **34**, 8780–8794 (2021)
45. Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M.. Lerchner, A.: beta-VAE: Learning basic visual concepts with a constrained variational framework. International conference on learning representations. (2017). Retrieved from https://openreview.net/forum?id=Sy2fzU9gl

46. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., others: An image is worth 16x16 words: Transformers for image recognition at scale. (2020). Preprint at http://arxiv.org/abs/2010.11929
47. Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., others: Dinov2: Learning robust visual features without supervision. (2023). Preprint at arXiv:2304.07193
48. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural. Inf. Process Syst. 25 (2012)
49. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L., C.: Mobilenetv 2: inverted residuals and linear bottlenecks. Proceedings of the ieee conference on computer vision and pattern recognition 4510–4520 (2018)
50. Tan, M., Le, Q.: Efficientnetv2: smaller models and faster training. International conference on machine learning 10096–10106 (2021)
51. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. Proceedings of the ieee conference on computer vision and pattern recognition 1492–1500 (2017)
52. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. (2014). Preprint at arXiv:1409.1556
53. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. Proceedings of the ieee conference on computer vision and pattern recognition 4700–4708 (2017)
54. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., others: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. (2021). Preprint at http://arxiv.org/abs/2107.02314

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

Authors and Affiliations

M. Mücahit Enes Yurtsever[1] · Süleyman Eken[1] · Bilgehan Arslan[2] · Yılmaz Atay[2] · Şeref Sağıroğlu[2] · Emrah Celtikci[3] · Alp Ozgun Borcek[3] · Alperen Sozer[4] · M. Caglar Sahin[5] · Zuhal Demirtas[6] · Kerem Nernekli[7]

✉ Süleyman Eken
suleyman.eken@kocaeli.edu.tr

M. Mücahit Enes Yurtsever
enes.yurtsever@kocaeli.edu.tr

Bilgehan Arslan
bilgehanarslan@gazi.edu.tr

Yılmaz Atay
yilmazatay@gazi.edu.tr

Şeref Sağıroğlu
ss@gazi.edu.tr

Emrah Celtikci
emrahceltikci@gazi.edu.tr

Alp Ozgun Borcek
alpborcek@gazi.edu.tr

Alperen Sozer
alperen.sozer@saglik.gov.tr

M. Caglar Sahin
m.sahin3@saglik.gov.tr

Zuhal Demirtas
zuhal.demirtas1@saglik.gov.tr

Kerem Nernekli
kerem.nernekli@mssm.edu

[1] Information Systems Engineering, Kocaeli University, 41001 Kocaeli, Türkiye

[2] Computer Engineering, Gazi University, 06570 Ankara, Türkiye

[3] Medicine, Gazi University, 06500 Ankara, Türkiye

[4] Neurosurgery, Sincan Training and Research Hospital, 06949 Ankara, Türkiye

[5] Neurosurgery, Kulu State Hospital, 42780 Konya, Türkiye

[6] Emergency Medicine, Alacam State Hospital, 55800 Samsun, Türkiye

[7] Internal Medicine, Icahn School of Medicine, 10029 New York, USA