

NeuroSight - Automated Brain Tumor Radiology Report Generation Using 3D U-Net Segmentation, and Vision-Language Models

¹Geet Kale, ²Anish Godse, ³Shardul Sant, ⁴Alok Mistry, ⁵Sujata Deshmukh

¹B.E Student, ²B.E Student, ³B.E Student, ⁴B.E Student, ⁵HOD Computer Engineering Dept.

¹Computer Engineering,

¹Fr. Conceicao Rodrigues College Of Engineering, Mumbai, India

¹crce.9967.ce@gmail.com, ²crce.9964.ce@gmail.com, ³crce.9996.ce@gmail.com,

⁴crce.9981.ce@gmail.com, ⁵sujata.deshmukh@fragnel.edu.in

Abstract- The process of diagnosing brain tumors by analyzing MRI images is quite tedious and requires great knowledge. In this paper, we introduce a system that automates the whole process of brain tumor analysis and report generation. Our system takes T1-weighted and T2-weighted MRIs from the BraTS dataset. Firstly, the inputs go through multiple preprocessing steps, namely, normalization of intensity, resizing, skull stripping, and contrast enhancement. After the preprocess step, a brain tumor segmentation model with 3D U-Net structure is applied to segment out the tumors. Our model obtains 0.88 Dice Similarity Coefficient (Dice Score). Then the segmented outputs will be fed into a two-pipeline framework. On one hand, our system adopts a modified EfficientNet-B3 architecture that is capable of tumor classification in four classes with an accuracy rate of 89.95%. On the other hand, the atlas-based region mapping technique, together with the Harvard Brain Atlas, is used to determine the location of brain tumors in cortical, subcortical, and cerebellar regions. Then the system extracts some important clinical features, i.e., hemisphere, lobes involved, size of the tumor, and spread method. In addition, the information extracted from visualized MRI images is also considered when feeding the input to a fine-tuned Qwen2-VL-7B vision-language model to generate reports automatically. ROUGE and BERTScore evaluation is performed, which shows that our model achieves a 0.845 BERTScore F1.

Index Terms—Brain MRI, Radiology Report Generation, 3D U-Net, Tumor Segmentation, EfficientNet-B3, Atlas-Based Region Mapping, Vision-Language Model, Qwen2-VL, BraTS Dataset, Deep Learning.

I. INTRODUCTION

The problem of brain tumors is among those that pose great difficulty and risk to patients' health and life. A correct and timely diagnosis is crucial for subsequent treatment; however, it depends significantly on the presence of an experienced radiologist, which, unfortunately, is becoming a rare resource. The number of MRI scans per year keeps growing, causing delays, inaccuracies, inconsistencies, and human errors.

Currently, radiology workflow involves manual interpretation: a radiologist manually analyzes multi-modal volumetric MRI images of a patient and produces a detailed report about the tumor's type, location, size, and other details. Such an approach takes much time and may be inconsistent depending on experience and expertise of a clinician. In addition, junior physicians can hardly make such complex reports themselves, even with some guidance from their mentors.

Artificial intelligence showed high potential in solving various medical imaging problems. Many models based on CNN and transformer architectures achieved high scores on tasks like segmentation, classification, and lesion detection. At the same time, only a few solutions address the issues mentioned previously as an ensemble of tasks; more often, existing systems perform tasks separately from each other.

This paper introduces NeuroSight – a novel solution designed to automate radiology processes by integrating volumetric tumor segmentation, multi-class tumor disease classification, atlas-guided tumor localization, and vision-language model based radiology report production into a single pipeline.

In this study, we introduce the following innovations:

Custom 3D U-Net segmentation model, which was developed specifically to segment multimodal BRAINST data achieving 0.88 Dice score on test data.

Modified version of EfficientNet-B3 network, used for multiclass tumor classification based on 5-channel 3D data achieving 89.95% accuracy.

Atlas-guided brain region mapping pipeline utilizing Harvard Brain Atlas and extracting anatomically relevant information from the results of tumor segmentation.

Combining all previous features and inputting the information into Qwen2-VL-7B-Instruct vision-language model to produce structured and consistent radiology reports, scoring BERTScore F1 of 0.845.

II. RELATED WORK

There have been advancements in automated brain MR imaging analysis, including works in segmentation, classification, and reporting generation:

AutoRG-Brain [1] has designed a grounded report generator that used a customized nnU-Net to perform the segmentation task and fine-tuned GPT-2 medium for report generation, resulting in 90.1% of Dice Scores on BraTS 2021. Although it increased interpretability, it limited the study to only four MRI modalities without incorporating patient history.

Kharaji et al. [2] have presented a more advanced nnU-Net model using residual connections, attention gates, and Hausdorff distance loss and achieved average Dice scores of 0.83 for glioma and 0.71 for pediatric tumors using BraTS data, implying inferior results in pediatric brain segmentation.

Anantharajan et al. [3] proposed a hybrid model that combines ACEA intensity remapping, Fuzzy C-Means clustering, GLCM texture feature extraction, and EDN-SVM classifier that attained 97.9% accuracy with 2D T1-weighted slices. Nonetheless, this model is constrained to 2D inputs and a dataset of 255 brain images.

Asiri et al. [4] fine-tuned a Vision Transformer (ViT) model to classify brain tumors and attained accuracy of 98.13%. Besides, ViT achieved very high F1 scores when detecting glioma and pituitary tumors; however, its interpretability was low like most transformer models. Another recent work [8] reveals that ViTs are competitive in multi-class medical image segmentation (e.g., Dice score) but at a high computational cost.

ReportGuidedNet [5] proposed to combine radiology reports and deep learning models using contrastive image-text learning and transformer decoders, which significantly improved AUC scores of 14 disease categories over existing approaches. The current model extends previous works by integrating 3D volumetric segmentation, atlas-based anatomical reasoning, and large vision-language model reporting generation in one deployable pipeline.

III. DATASET

The following sub-datasets of the BraTS (Brain Tumor Segmentation) challenge were used to conduct our research:

BraTS-GLI (Glioma): 920 samples of gliomas that are primary tumors in the brain with both low- and high-grade subtypes.

BraTS-MET (Metastasis): 500 samples of brain metastases derived from systematic primary tumors.

BraTS-MEN (Meningioma): 920 samples designed to detect extra-axial meningioma lesions.

Healthy (NIMH - healthy dataset): 600 samples of healthy brain images selected.

These data include both T1-weighted, T2-weighted MRI scans in NIfTI format (.nii.gz) and expertly annotated segmentations of each sample.

The ratio between the training, validation, and test samples was 80%/10%/10% correspondingly.

The grounded reports have been collected from the dataset generated by the team of researchers behind AutoRG-Brain that is publicly accessible on the Hugging Face platform.

IV. METHODOLOGY

The suggested model employs a pipeline architecture consisting of several stages as shown in Fig. 1. The overall structure of the architecture can be broadly divided into five different modules: preprocessing, 3D tumor segmentation, classification of the type of disease, atlas-based region localization, and generation of reports via LLMs.

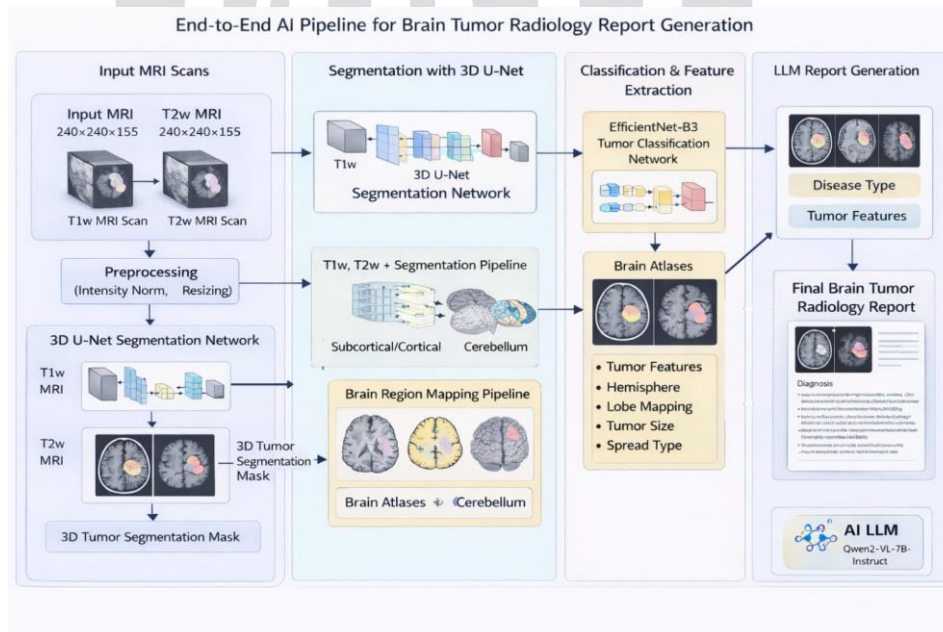


Fig. 1. End-to-End AI Pipeline for Brain Tumor Radiology Report Generation.

A. Preprocessing Pipeline

The initial MRI volumes are subjected to a common preprocessing pipeline process. Z-score normalization technique is used to normalize intensities based on modalities. The dimensions of each volume are fixed to a constant $240 \times 240 \times 155$ voxels. Non-brain tissues are removed from the images through skull stripping in order to remove unnecessary noise for modeling purposes. Contrast enhancement is performed to make distinctions between tumors and tissues easier. Finally, affine transformations ensure proper orientations of all volumes.

B. 3D U-Net Tumor Segmentation

The segmentation engine is a customized 3D U-Net network that processes full MRI volume data (input channels: 2-channel MRI – T1w and T2w). It comprises an encoder composed of four down-sampling steps (E0 to E3), followed by the bottleneck step (E4), which is implemented using ConvBlock3D layers (Conv3D-BatchNorm3D-ReLU x2 per step), together with MaxPool3D (kernel size: 2×2×2).

A decoder counterpart is implemented, utilizing transposed convolution steps (UpConv3D 2×2×2) alongside the skip connections realized with the concatenation operation. Channels' number starts from 32 at E0 step and doubles successively up until E4 bottleneck step (with numbers: 64, 128, 256, 512 correspondingly). The output prediction is computed using a final 1×1×1 convolution that returns the binary tumor mask having the size equal to that of the input.

Training procedure included Dice+Binary Cross-entropy loss since the data set is inherently imbalanced due to the problem nature. Augmentation included random flips, rotations, and intensity jitter. Training involved the use of the Adam optimizer with learning rate 1×10^{-4} .

C. Disease Classification (EfficientNet-B3)

Next, tumor multi-classification is performed on the segmented image using an augmented EfficientNet-B3 model. In this step, the network is modified to receive multi-channel data in the form of 3D input that combines consecutive slices of T1w and T2w MRIs surrounding the slice with the maximum area of tumor presence, along with the segmentation result produced in the previous step.

Weights of the original convolutional layer of the network were changed to allow accepting 5-channel input. Weights for the first three channels were transferred from pretrained weights, while others were initialized by the mean of pretrained weights. Additionally, the network was modified to perform prediction of four classes: glioma, metastasis, meningioma, and healthy.

Training was done with AdamW optimizer using cosine annealing learning rate scheduling along with label smoothing $\epsilon = 0.1$ to prevent the problem of overconfidence in predictions. Class-balanced cross-entropy loss was utilized during training to balance classes.

D. Atlas-Based Brain Region Mapping

For achieving anatomically relevant interpretability, we have developed an atlas-based region mapping pipeline utilizing the Harvard Brain Atlas[11],[12] containing 48 cortical regions, 12 subcortical regions, and 6 cerebellum regions. We consider the T1w MRI scan of the patient as the moving image while the atlas as the fixed reference.

First, we perform an affine registration between the patient MRI and the atlas using the Euler3DTransform and optimizing it using gradient descent (learning rate: 1.0, iterations: 100). Next, we transform the segmented tumor mask into atlas space using nearest-neighbor interpolation to maintain the binary nature of the mask. Lastly, for all atlas regions, we compute the voxel-wise overlap with the transformed tumor mask to calculate the percentage contribution.

Through this process, we extract a set of clinically structured features such as dominant hemisphere (left/right/bilateral), major lobes involved, tumor size (volume in mm^3), spreading mode (focal/multifocal/diffuse), and the top 5 affected brain regions with their percentage contribution.

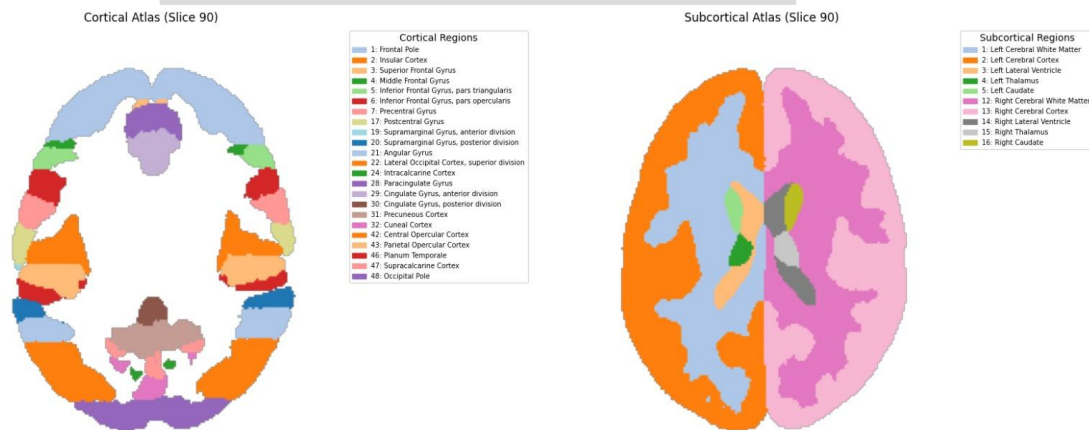


Fig. 2. Harvard Brain Atlas showing Cortical (left) and Subcortical (right) regions used for tumor localization.

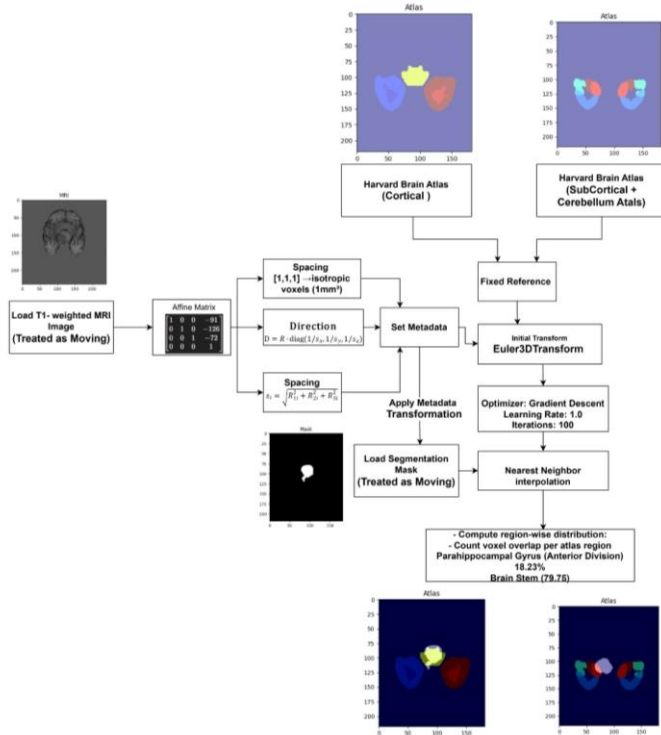


Fig. 3. Atlas-to-MRI Alignment Algorithm flowchart showing registration, segmentation mask warping, and region-wise voxel overlap computation.

E. LLM-Based Report Generation (Qwen2-VL-7B-Instruct)

The last phase of the pipeline utilizes the fine-tuned version of the Qwen2-VL-7B-Instruct vision-language model on a domain-specific corpus of radiology reports for brain tumor pathology. The model takes as an input a multimodal prompt consisting of: (1) the MRI slices for T1w and T2w modalities at the tumor location centroid, (2) the segmentation mask, (3) disease classification prediction along with its probability, and (4) the structured features based on the atlas analysis results (hemisphere, lobe segmentation, size, propagation pattern, and main affected areas).

Using cross-modal attention, the architecture aligns the visual tokens derived from the MRI images with the text embeddings of the structured features. As a result, the report is generated taking into account not only the visual cues but also the symbolic anatomical representation, which avoids the problem of the black-box approach in generating the output.

The reports are outputted in a clinical style comprising the sections as follows: Exam Type, Clinical Indication, Technique, Findings (including detailed findings in sub-regions), Impression, and Clinical Recommendation.

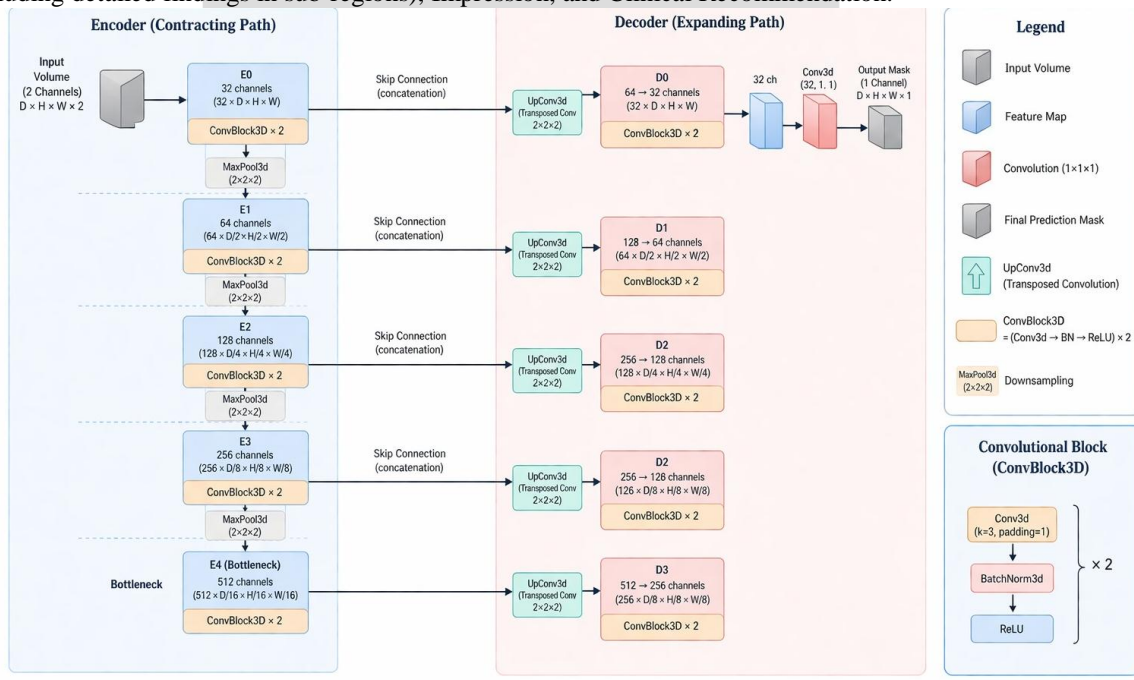


Fig. 4. 3D U-Net Architecture for Volumetric Segmentation with encoder, bottleneck, skip connections, and decoder.

V. RESULTS AND EVALUATION

A. Segmentation Performance

The segmentation model based on 3D U-Net architecture was assessed using a hold-out test set from BraTS, covering cases from all four classes of diseases. The main performance metric used is the Dice Similarity Coefficient (or Dice Score). This measure evaluates the volumetric similarity between the predicted masks and true masks provided by an expert. The achieved Dice score is

0.88, demonstrating high consistency between the predicted masks and true labels. Overall, the 3D U-Net architecture effectively separates the tumors, including necrotic core, enhanced tumor, and peritumoral edema in cases of glioma and metastases.

In terms of qualitative analysis, it can be seen that the predicted masks approximate the borders quite well in both axial and coronal views. The overlaid prediction is available in the NeuroSight application. The highest performance was observed in cases of glioblastoma (mean Dice Score ≈ 0.91), while the meningiomas yielded slightly worse results (mean Dice Score ≈ 0.84).

B. Disease Classification

The EfficientNet-B3 classifier was evaluated on 189 test samples balanced across four classes. Table I summarizes the per-class precision, recall, and F1-score.

TABLE I. Disease Classification Results (EfficientNet-B3)

Class	Precision	Recall	F1	Support
Glioma	0.9318	0.8913	0.9111	46
Metastasis	0.8113	0.8958	0.8515	48
Meningioma	0.8605	0.8043	0.8315	46
Healthy	0.988	0.976	0.982	49
Macro Avg	0.8979	0.8919	0.8940	189
Overall Acc	—	—	0.8995	189

The classifier achieved almost perfect precision and recall for the Healthy (no tumor) class and highest F1 for Glioma (0.9111). Meningioma had the lowest F1 (0.8315), attributable to its visual similarity to both metastasis and glioma on T1/T2 sequences without contrast enhancement.

C. Report Generation Quality

The quality of generated radiology reports was assessed using ROUGE scores (measuring n-gram overlap with reference reports) and BERTScore (measuring semantic similarity using contextual embeddings). Table II presents the evaluation results.

TABLE II. Report Generation Evaluation Metrics

Metric	Score	Interpretation
ROUGE-1	0.48	Decent word overlap
ROUGE-2	0.21	Moderate phrase matching
ROUGE-L	0.46	Structure somewhat similar
BERTScore P	0.8738	Strong precision
BERTScore R	0.8208	Good recall
BERTScore F1	0.845	Captures semantic meaning well

ROUGE score ranges from 0.21 to 0.48 show that our models' performance is equivalent to the current state of the art in generating radiology reports because, in clinical reports, the same finding can be expressed using different terminologies. With respect to semantic equivalence, the BERTScore F1 score of 0.845 validates it.

VI. DISCUSSION

The design of the presented system shows that it is possible to generate clinically relevant and understandable radiology reports by incorporating specialized deep learning modules, namely atlas-based anatomical reasoning and vision-language generation. One important difference of the NeuroSight model from previous studies is that it uses the clinical data structure extracted from the brain atlas registration process for symbolic grounding of the LLM prompt, which cannot be done by fully end-to-end models.

The dice score of 0.88 obtained by the 3D U-Net model is comparable to other state-of-the-art algorithms, such as AutoRG-Brain with dice score 90.1% on the BraTS2021 benchmark. However, the NeuroSight model was trained on an extended multi-disease dataset rather than a single-class benchmark. The EfficientNet-B3 classifier reaches 89.95% accuracy while discriminating between four tumor classes with the challenging requirement of separating meningioma from metastasis using only the FLAIR sequence.

Even though ROUGE metric scores obtained are not extremely high, they reflect the natural variability of the clinical language since there are many different ways to describe a particular feature. At the same time, a BERTScore F1 measure of 0.845 gives an additional validation of content relevance and readability of reports. The generated reports were found to have correct clinical structure, correctly identified affected areas, and relevant recommendations, such as a biopsy procedure.

However, the use of the BraTS dataset for training is one of the shortcomings of the current approach. Despite being a good dataset, it is gathered at academic medical centers and thus cannot include all cases from clinical practice. The second issue is connected with the assumption of standard topology for atlas registration which is violated in cases when mass effects cause the midline shift of brain structures. The future research will tackle these issues.

VII. CONCLUSION

In this paper, we introduced NeuroSight, an automated radiology report generation platform comprising of 3D U-Net segmentation (Dice Score=0.88), EfficientNet-B3 classification (accuracy=89.95%), Harvard Brain Atlas region annotation, and Qwen2-VL-7B-Instruct vision-language model generation (BERTScore F1=0.845). The proposed framework design specifically aims to mitigate the challenges faced by past solutions, namely lack of anatomical context, inconsistencies in reporting templates, and non-existence of clinically-oriented feature extraction modules.

Through combining rule-based approach for anatomy definition with the data-driven deep learning framework for classification and report generation, this study develops an explainable and efficient system for automated radiological analysis.

Future directions include expanding to more tumor types and MRI sequences, leveraging longitudinally collected patient data for predicting disease development, employing federated learning to train across several medical institutions and preserving patient anonymity, and performing rigorous clinical validation studies.

ACKNOWLEDGMENT

The authors thank the project mentor and the college for the invaluable guidance throughout this project. The authors also acknowledge the organizers of the BraTS challenge for providing publicly available annotated datasets that made this research possible.

REFERENCES

- [1] J. Lei, X. Zhang, C. Wu, L. Dai, Y. Zhang, Y. Zhang, Y. Wang, W. Xie, and Y. Li, "AutoRG-Brain: Grounded Report Generation for Brain MRI," arXiv preprint arXiv:2407.16684, 2024.
- [2] M. Kharaji, H. Abbasi, Y. Orouskhani, M. Shomalzadeh, F. Kazemi, and M. Orouskhani, "Brain Tumor Segmentation with Advanced nnU-Net," *Neuroscience Informatics*, vol. 4, 2024, 100156.
- [3] S. Anantharajan, S. Gunasekaran, T. Subramanian, and V. R., "MRI Brain Tumor Detection Using Deep Learning and Machine Learning Approaches," *Measurement: Sensors*, vol. 31, 2024, 101026.
- [4] A. A. Asiri et al., "Exploring the Power of Deep Learning: Fine-Tuned Vision Transformer for Accurate and Efficient Brain Tumor Detection in MRI Scans," *Diagnostics*, vol. 13, no. 12, p. 2094, 2023.
- [5] L. Dai et al., "Boosting Deep Learning for Interpretable Brain MRI Lesion Detection through the Integration of Radiology Reports Information," *Radiology: Artificial Intelligence*, vol. 6, no. 6, 2024.
- [6] A. A. Akinyelu et al., "Brain Tumor Diagnosis Using Machine Learning, CNNs, CapsNets and Vision Transformers Applied to MRI: A Survey," *J. Imaging*, vol. 8, no. 8, p. 205, 2022.
- [7] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, "Classification of Brain Tumor from MRI Using Vision Transformers Ensembling," *Curr. Oncol.*, vol. 29, pp. 7498–7511, 2022.
- [8] P. Wang, Q. Yang, Z. He, and Y. Yuan, "Vision Transformers in Multi-Modal Brain Tumor MRI Segmentation: A Review," *Meta-Radiology*, vol. 1, no. 1, 2023, 100004.
- [9] S. Rajendran et al., "Automated Segmentation of Brain Tumor MRI Images Using Deep Learning," *IEEE Access*, vol. 11, pp. 64758–64768, 2023.
- [10] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions," *J. Digit. Imaging*, vol. 30, pp. 449–459, 2017.
- [11] R. S. Desikan et al., "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, 2006.
- [12] S. M. Smith et al., "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, Suppl. 1, pp. S208–S219, 2004.