Enhancing Brain Tumor Detection in Magnetic Resonance Imaging Through Explainable Artificial Intelligence Techniques and Fusion Models

Adwaita Sathrukkan[1][a], Naveen Raaghavendran[1][b], Sanjay Balamurugan[1][c], Srivaishnavi J. V.[1][d], S. Raghul[2][e] and G. Jeyakumar[1][f]

[1]*Department of Computer Science Engineering, Amrita School of Computing, Coimbatore, Amrita Vishwa Vidyapeetham, India*

[2]*Research and Engineering Scientist, Zoho Corporation Private Limited, India*

Keywords: Image Recognition, Deep Learning, Convolutional Neural Networks, Tumor Detection, Explainable Artificial Intelligence, Feature Extraction.

Abstract: Enhancing the detection of brain tumors in Magnetic Resonance Imaging (MRI) represents a critical frontier in medical imaging and neuro-oncology. This paper introduces an innovative approach that leverages Explainable Artificial Intelligence (XAI) techniques and fusion models to significantly improve the accuracy and interpretability of brain tumor detection. This paper proposes a novel framework integrating deep learning models and fusion strategies for enhanced feature extraction from multiple MRI sequences, as detailed in subsequent sections. By employing XAI methodologies, the approach presented in this paper not only enhances detection performance but also provides meaningful explanations for its predictions, thereby increasing the trustworthiness of automated diagnosis.

1 INTRODUCTION

Tumors of the brain and other parts of the nervous system, such as glioblastomas (GBM), are among the top causes of cancer mortality in adult populations. Brain tumors, whether malignant or non-malignant, constitute the second-highest cause of death linked to cancer in adolescents and children. Standard treatments for brain cancer encompass surgery, radiation therapy, and chemotherapy. However, surgically pinpointing and removing the diseased areas is often exceedingly challenging due to the complexity involved in distinguishing tumors from the normal brain tissue visually. Magnetic resonance imaging (MRI) is a crucial tool in clinical settings, aiding doctors brain tumor identification. MRI provides detailed images of soft tissues, which enhances the ability to determine the location and boundaries of tumors.

The advent of machine learning (ML) models in medical imaging marks a significant leap forward in diagnostics, particularly in the domain of brain tumors. These advanced computational tools have demonstrated the ability to analyze complex imaging data with high precision, offering insights into tumor characteristics that were previously unattainable through traditional diagnostic methods. Despite these advancements the integration of ML models into clinical practice faces considerable challenges, primarily due to "Black Box" nature of Artificial Intelligence (AI) algorithms. This opacity in decision-making processes poses a barrier to clinical adoption,

[a] https://orcid.org/0009-0007-0509-8659
[b] https://orcid.org/0009-0006-6253-558X
[c] https://orcid.org/0009-0007-9025-9532
[d] https://orcid.org/0009-0009-3993-0794
[e] https://orcid.org/0000-0002-1306-6960
[f] https://orcid.org/0009-0009-3993-0794

Sathrukkan, A., Jeyakumar, G., Balamurugan, S., J. V., S., Raghul, S. and Raaghavendran, N.
Enhancing Brain Tumor Detection in Magnetic Resonance Imaging Through Explainable Artificial Intelligence Techniques and Fusion Models.
DOI: 10.5220/0013344000004646

In *Proceedings of the 1st International Conference on Cognitive & Cloud Computing (IC3Com 2024)*, pages 266-275
ISBN: 978-989-758-739-9

as healthcare professionals require transparent and interpretable systems to trust and effectively use these technologies in patient care.

Explainable Artificial Intelligence (XAI) emerges as a crucial area of research aimed at addressing these challenges. By making AI's decision making transparent and understandable, XAI holds the promise of bridging the gap between the technical capabilities of ML models and the practical need of clinical diagnosis. However, a comprehensive literature survey reveals a significant lack of focused research on the application of XAI in the domain of medical imaging for brain tumor detection. While several studies have underscored the potential of ML in improving diagnostic accuracy, the exploration of XAI models in elucidating the rationale behind AI-generated diagnoses remains limited. This gap in research underscores the significance of this study, which seeks to investigate the integration of XAI within the context of brain tumor detection from medical images. The work presented in this paper aims to not only highlight the potential of combining ML models with XAI techniques to enhance diagnostic accuracy but also to address the pressing need for interpretability in medical AI applications. By doing so, this study contributes to the broader adoption of AI in healthcare, ensuring that the benefits of these technologies can be fully leveraged to improve patient outcomes while maintaining the trust and confidence of medical practitioners in AI-driven diagnostic tools.

The remaining part of the paper is organized as follows. Section 2 explains various state-of-the- art ML and XAI models. The Section 3 briefs about various ML models used in medical image processing. Section 4 briefs about the usage of XAI models in medical image processing. Next, section 5 presents the details of the proposed methodology. Section 6 presents the design details of the experiments. The results and discussions are in Section 7. The Section 8 concludes the paper.

2 LITERATURE SURVEY

The initial phase of medical imaging involves the detection of tumors in the MRI scans and subsequent extractions of essential features for classifications as presented in (Abhilasha et al., 2022). Numerous methodologies have been developed to address challenges associated with variations in field strength, dataset biases, mislabeled in- stances, and other illustrative changes in the context of medical imaging. The evolution from conventional hand-written medical diagnosis, by the people in the field, to deep learning-based models and AI have proven advantageous, particularly in handling large data and providing a robust feature representation, segmentation, and classification. In the do- main of XAI, innovative approaches of using XAI in deep learning-based medical image analysis are de- scribed in (Velden et al., 2022). As given in (Priya, A. and V. Vasudevan, 2024), brain tumor classification and detection are possible using a suitable CNN structure (Eg. a hybrid AlexNet-GRU) based on the given MRI data. This process involves sharpening and denoising the MRI images using local filters.

A feature extraction method from brain MRIs is proposed in (Tas, 2023), and is used for brain tumor detection. In this work the DenseNet201 is trained using the exemplar method, and then the features are extracted. The authors of (Amran et al., 2022), proposes a brain tumor classification and detection system using a GoogLeNet architecture. In the proposed architecture, 5 layers of GoogLeNet are eliminated and 14 new layers are added to extract the features automatically. In (Apostolopoulos et al., 2023), a novel approach of integrating CNN with attention models and feature-fusion blocks is presented. This integrated approach is demonstrated on the brain tumor classification task, using MRI data. This approach is named as Attention Feature Fusion VGG19 (AFF- VGG19), and it was found outperforming other state- of-the-art similar approaches. Regarding the basis of segmentation, a brain tumor segmentation system using modified ResUNET architecture which combines the strengths of the U-Net architecture is presented in (Pathak et al., 2023). This system is known for its effectiveness in bio-medical image segmentation, with the residual learning framework to facilitate training of deeper networks as evidenced in ((Pathak et al., 2023).

The work presented in (Younis et al., 2022), explored the usage of VGG (Visual Geometry Group) and CNN for brain tumor related image analysis. A new system is proposed and the same is demonstrated for training and classifying brain tumors based on different MRI images. Also, it is found from the literature that different CNNs such as VGG16/19, AlexNet, GoogLeNet and Resnet are demonstrating well on MRI based image classification tasks. In the field of XAI, automatic segmentation of multimodal brain tumor images based on classification of super- voxels which uses a type of MRI sequence called Fluid Attenuated Inversion Recovery (FLAIR) is popular. FLAIR is an imaging technique used to sup- press the effects of fluid within the image, particularly cerebrospinal fluid (CSF), to bring out the periventricular hyperintensities (lesions near the ventricles of the brain), as give in the study (Hu et al., 2021). NeuroXAI uses seven advanced methods to clarify deep neural networks in MRI brain tumor analysis, providing visualization maps for transparency. These include: Vanilla Gradient (VG) for highlighting crucial image areas, Guided Backpropagation (GBP) for alternative gradient calculations, Integrated Gradients (IG) to tackle gradient saturation, Guided Integrated Gradients (GIG) for refined attribution paths, SmoothGrad for sharper sensitivity maps, Gradient CAM (GCAM) for model-agnostic visual explanations, and Guided GCAM (GGCAM) for high-resolution detail capture, as evidenced in (Zeineldin et al., 2022). A multi-disease diagnosis model using the X-ray images of the chest, with XAI, is presented in (Rani et al., 2022).

In-Hospital mortality prognosis, the usage of XAI techniques is demonstrated using seven different machine learning models in (Maheswari et al., 2023). A model for classifying suprasellar lesions formed in the brain is proposed in (Priyanka et al., 2023). This study has used discharge summary of 422 patients. The usage of machine learning models in diagnosis other medical issues also notable. As an example, a diabetic retinopathy detection using Gradient-weighted class activation map (Grad-CAM), presented in (Duvvuri et al., 2022), is added here. This literature survey summarizes advancements in brain tumor detection and classification through medical imaging, tracing the shift from traditional diagnostic methods to deep learning and AI technologies. It discusses challenges such as imaging variability and dataset biases, and the evolution towards automated feature extraction and classification using deep learning architectures like Dense-Net, GoogLeNet, and VGG-16. Highlighting the role of XAI in making neural network decisions transparent, the survey underscores the necessity for methods like Vanilla Gradient, Guided Backpropagation, and Gradient CAM to ensure model reliability and acceptance by medical professionals. Building on this, this paper proposes a novel approach that merges the latest in deep learning with XAI to enhance diagnostic accuracy and interpretability, aiming to revolutionize AI-driven medical imaging for brain tumor analysis.

3 MACHINE LEARNING MODEL

In the adapted implementation of the AlexNet architecture for experimental purposes, the model features five convolutional layers with kernel sizes 11x11 for the first layer and 3x3 for subsequent layers, followed by three maxpooling layers and enhanced with batch normalization to improve training efficiency. The architecture includes four dense layers with a substantial number of neurons (4096 for the first two dense layers, 1000 for the third) and employs dropout with a rate of 0.4 after each dense layer for regularization. The GoogLeNet architecture is utilized for brain tumor classification from MRI images, featuring multiple inceptions blocks that parallelly process input through convolutional layers of varying kernel sizes and a max-pooling layer, enhancing feature extraction efficiently. This implementation starts with a 7x7 convolution, progresses through strategic inception blocks and max pooling for depth and dimensionality reduction, and concludes with global average pooling and a SoftMax classification layer.

This GoogLeNet model, which is optimized with Adam and has Early Stop-ping and Model Checkpoint callbacks, is designed for high accuracy in multi-class classification tasks, indicating the potential of deep learning in medical diagnostics. For efficient MRI brain tumor segmentation, the VGG19+UNet architecture combines the reliable feature extraction of VGG19 with the accurate localization of UNet. With VGG19 pre-trained on ImageNet for deep feature extraction, this model performs exceptionally well at identifying complex patterns in MRI pictures. The UNet decoder uses these features for reconstructing the segmentation maps and to identify tumors. There is a preprocessing, augmentation, and normalization procedure applied to the MRI pictures and the segmentation masks before the training. The performance metrics and the visual evaluations prove that this hybrid approach shows high precision in tumor delineation.

ResNet's deep feature extraction procedure is merged with UNet's accurate localization. This merging uses ResNet's residual connections, prevents the vanishing gradient problem, and enhances learning efficiency. This approach used a custom data generator for data handling. The MRI images and masks are processed by resizing and normalization. The dynamic learning rate adjustments and early stopping through callbacks are the important training techniques used. The ability of this model to successfully segment brain tumors highlights the usefulness of integrating residual learning with UNet's architecture and highlights the model's potential for improving clinical diagnostics and medical imaging research. The modified VGG16 architecture, which uses 3x3 filters with ReLU activation, keeps its basic structure of 13 convolutional layers arranged into five blocks, each of which ends with max pooling to reduce dimensions and classify brain tumors from MRI images. A customized classifier, consisting of a flatten layer, dropout for regularization, and dense layers culminating in a SoftMax activation for multiclass prediction, replaces the original fully connected layers. By making the final convolutional block train- able to capture tumor-specific features, the model is refined.

4 XAI MODELS

To make the decision-making of neural networks transparent, the NeuroXAI framework combines an explanation generation module with a deep neural network for processing MRI brain scans. The process starts with MRI scans that are run through a convolutional neural network (CNN), which produces feature maps and outcomes such as tumor segmentations or classifications. Medical experts review these results, and upon request, the system employs advanced XAI techniques to generate visual explanation maps. These techniques have Vanilla Gradient (VG) and Gradient CAM (Grad-CAM). The VG is to create a saliency map and to identify influential image parts. The Grad-CAM is to highlight important regions for predictions. Also, the techniques of Integrated Gradients (IG) and Guided Backpropagation (GBP) are included in the framework. They are used to identify which areas of the images have a major impact on the decisions. The Guided Integrated Gradients (GIG) are used to further improve the feature relevancy attribution. The GIG uses the computation of gradients between the baseline and the input image. Smooth Grad and Guided Grad-CAM techniques average gradients of noise-perturbed input images and combine macro and micro-level visualizations, respectively, offering clearer, more interpretable visualizations. These sophisticated XAI approaches help bridge the gap between AI outputs and clinical decision-making, fostering trust and collaboration in AI-assisted diagnostics.

5 PROPOSED METHODOLOGIES

To address the challenge of MRI brain tumor classification and segmentation, this paper integrates advanced adaptations of deep learning architectures. Initially, a comprehensive dataset of MRI images is compiled and subjected to meticulous preprocessing, including normalization, resizing, and augmentation, to prepare for model training. This paper adapts and optimizes several renowned architectures for specific tasks: AlexNet is tailored for binary classification with adjusted convolutional layers and dropout rates; GoogLeNet is configured with inception blocks for efficient multi-class classification; VGG19 is combined with UNet for precise tumor segmentation through deep feature extraction and localization; ResUNet leverages ResNet's residual connections with UNet's segmentation accuracy; and VGG16 is modified with custom classifiers for enhanced tumor feature recognition. Each model undergoes fine-tuning, employing strategies like dynamic learning rate adjustments and early stopping, to ensure optimal performance.

This comprehensive approach, focusing on the customization of CNNs, aims to enhance the accuracy, efficiency, and reliability of MRI brain tumor diagnosis and segmentation, demonstrating the potential of deep learning in medical diagnostics and imaging analysis. Additionally, this paper integrates Explainable Artificial Intelligence (XAI) methods to enhance model transparency and interpretability. Techniques such as Vanilla Gradient, Grad-CAM, Guided Backpropagation, Integrated Gradients, Guided Integrated Gradients, Smooth Grad, and Guided Grad-CAM generate visual explanation maps, aiding medical experts in understanding AI decisions. The NeuroXAI framework processes MRI scans through a CNN, creating feature maps and results with clear visual explanations. Model performance is evaluated using metrics like Intersection over Union (IoU) and accuracy, comparing predictions with manual segmentations. Clinical validation through trials and feedback from medical professionals ensures real-world applicability and reliability. This approach highlights the potential of XAI techniques and fusion models to im- prove diagnostic accuracy, clinical decision-making, and patient outcomes in neuro-oncology.

6 DESIGNS OF EXPERIMENTS

Different datasets have been used during model selection and in the stage of implementation of the final XAI model. Brain MRI images with manual FLAIR abnormality along with segmentation masks are obtained from (Buda, Mateusz, 2022). This dataset was obtained from The Cancer Imaging Archive (TCIA). This dataset has the details of 110 patients. This dataset is used as a base dataset to find the best model among the 6 chosen models. The datasets used for training, validation and testing are available in (Shah, 2019). This includes pre-operative multimodal MRI scan of glioblastoma (HGG) and lower grade glioma (LGG). This BRATs 2019 dataset is used for Classification task and the image format of the dataset is 2-dimensional. BRATs 2021 dataset is used for segmentation task with 3-dimensional images obtained from (Schettler, 2021).

This study considers the VGG16, VGG19+unet, ResUnet, Alexnet, Googlenet and CNN ML Models. The machine learning algorithms are dependent on parameters, initial values and the training/testing process depends on updating the values until the requirements are met. Parameters are particularly important for fine tuning of the model, making proper predictions and defining the skill of the model on the given problem. Table 1 presents the information of all the parameters that are important to detect the anomaly in the brain MRI image and get the highly accurate machine learning models with the help of performance metrics. The experiments are done with SmoothGrad, Guided Grad-CAM, Guided Backpropagation, Vanilla Grad, Grad-CAM and Guided Integrated Gradients, which are the XAI Models.

Parameters in XAI models are important in shaping the interpretability and transparency of the model, to understand the decision-making process efficiently. The parameter sample size plays a key role here because it generates accurate heatmap in the output of all the 8 different XAI models and it defines the number of trainings also. Table 2 presents a detailed breakdown of the important parameters used in the models.

7 RESULTS AND DISCUSSIONS

Machine learning performance measures are essential determinants of model's efficacy and capacity to complete the assigned task. Precision, accuracy, and F1 score are the commonly used metrics in machine learning and data analysis for evaluating the performance of classification models. These metrics are used together to provide a comprehensive evaluation of the model's performance, considering distinct aspects of classification accuracy and error. Table 3 results help us to get the best model, which is a fusion of two base models namely, VGG19 and Unet. These results are used as a base model for the other 8 XAI models for generating heat maps to detect anomaly in the brain's MRI images.

Table 4 results in the performance metrics used in the image segmentation process. The metric Intersection over Union (IoU) and Monte Carlo prediction are specific for XAI methods. The ratio of the inter- section area between the predicted and the ground truth masks to the union of both the masks, is used as the measure for IoU. Monte Carlo predictions are often used for uncertainty prediction particularly for a Bayesian deep learning.

The mean prediction shape represents the average, or the expected value of the predictions generated by the segmentation process for a particular slice index. The mean prediction matrix represents the predicted values for each vowel in the input volume, in this case particularly it has taken intensities for each 4 of the classes.

In general, a machine learning model defines a representation of a training process. It does the task of discovering the patterns from the input training data, it structures a ML model which can understand these patterns and makes predictions on new inputs.

There are three types of learning algorithms that can be followed in a machine learning model - Supervised learning, Unsupervised learning, and Reinforcement learning. The chosen 6 classification models follow super-vised learning where the relationship between the input and output is designed, and it uses labeled datasets to train the algorithm to predict the outcomes and recognize patterns.

Table 1: Parameter configuration of chosen ML models.

S no.	Model Name	Epoch	Batch size	Train:Test:Valid	Total params
1	VGG16	7	32	2828:393:708	128590
2	VGG19+unet	2	36	1167:103:103	31172033
3	ResUnet	100	16	1167:103:103	1210513
4	Alexnet	150	17	2611:650:329	26052829
5	Googlenet	20	30	2611:653:326	5977692
6	CNN	35	32	2504:835:590	18818113

Table 2: Parameter Requirements for Various XAI Models.

Parameter Name	Smooth Grad	Guided Grad-CAM	Guided Backpropagation	Vanilla Grad	Grad-CAM	Guided Integrated Gradients
model	Required	Required	Required	Required	Required	Required
Io_imgs	Required	Required	Required	Required	Required	Required
Class_id	Required	Required	Required	Required	Required	Required
LAYER_NAME	Optional	Optional	Optional	Optional	Optional	Optional
MODALITY	Optional	Optional	Optional	Optional	Optional	Optional
XAI_MODE	Optional	Optional	Optional	Optional	Optional	Optional
DIMENSION	Optional	Optional	Optional	Optional	Optional	Optional
STDEV_SPREAD	Optional	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
N_SAMPLS	Optional	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
MAGNITUDE	Optional	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable
CLASS_IDs	Not applicable	Optional	Not applicable	Not applicable	Not applicable	Not applicable
TUMOR_LABEL	Not applicable	Optional	Not applicable	Not applicable	Not applicable	Not applicable
eps	Not applicable	Optional	Not applicable	Not applicable	Not applicable	Not applicable
STEPS	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Required
FRAC	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Required
MAX_DIST	Not applicable	Not applicable	Not applicable	Not applicable	Not applicable	Required

Table 3: Evaluation metrics for ML models.

S no.	Model Name	Precision	Accuracy	F1 score
1	VGG16	85	82.19	83
2	VGG19+unet	98.5	98.27	97
3	ResUnet	98	96	97
4	Alexnet	57	60	72
5	Googlenet	88	87	87
6	CNN	95.42	95.96	95.42

Table 4: Evaluation Metrics for XAI Methods.

S.No	Metrics Name	Values
1	Intersection over Union (IoU)	-1.035
2	Monte Carlo Prediction	Mean prediction shape (2,1,192,244,160,4)

Figure 1 shows a clear view of the fusion model's accuracy and its high performance. Selecting VGG19 and UNet as a combinational model is the base and the next step is to work with the XAI models and check the accuracy of them along with this machine learning model.

The two types of XAI methods are model-specific and model agnostic. Among these two, the former ones are tailored to unique characteristics and architectures of a particular machine learning model. These methods aim to provide explanations specifically designed for the internal working of the chosen model.

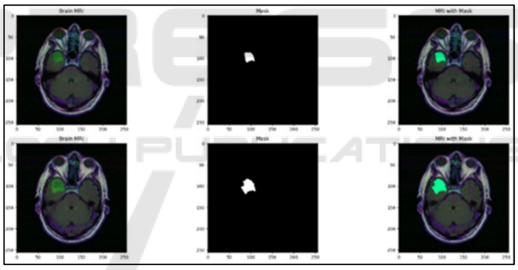

Figure 1: Accurate prediction of the tumor in the brain image using the mask component by the fusion model. A) Input MRI. B) Mask used. C) MRI with the mask.

Model-agnostic approaches, on the other hand, are designed to be versatile and applicable to various machine learning models. The chosen XAI methods are FLAIR, Vanilla, Back propagation, IG, Guided IG, Smooth Grad, Grad-CAM, Overlay Grad-CAM, Guided Grad-CAM, Prediction and Prediction-Overlay. These models fall under model-agnostic methods as they rely on computing the gradients and based on intricate details about the architecture. The difference among these methods lies in their specific techniques for highlighting and explaining distinct aspects of machine learning model's decision-making process. Tables 5 and 6 explain the functionality of each method and its contribution in classification and segmentation.

Table 5: Explanation of selected XAI methods – Part I.

S.No.	Method Name	Explanation Methodology	Application in Classification and Segmentation
1	FLAIR	Highlights notable features in an image.	Reveals the important feature that contributes significantly to the model decision.
2	Vanilla	Computed gradients of the output concerning the input pixels.	Useful in both the tasks and highlights areas where slight changes in pixel value occur.
3	Integrated Gradients (IG)	Computed the integral of gradients along the path.	Offers a holistic view of pixel importance.
4	Guided IG	Restricts the backpropagation of gradients.	Emphasizes positive influence on the model's decision.

Table 6: Explanation of selected XAI methods – Part II.

S.no	Method Name	Explanation Methodology and Application in Classification and Segmentation
1	Smooth Grad	Adding random noise to the input image and averaging the resulting gradients.
2	Grad-CAM(class activation map)	Generates a heatmap highlighting the regions where the model focused during the decision making.
3	Overlay Grad-CAM	This overlays the generated heatmap onto the original image for more intuitive visualizations. Effective in classification tasks.
4	Guided Grad-CAM	Combines the guided back-propagation with Grad-CAM and provides localization and guidance on which features contribute positively.
5	Prediction Overlay	Overlays the predicted class onto the input image and gives a direct image of the model's decision.

The implementation details of the selected XAI models are discussed under the following three categories:

(1) classification (based on VGG19 model)
(2) segmentation
(3) CNN segmentation (convolutional neural network)

7.1 Classification Based on VGG19 Model

The dataset classifies images as either indicating glioma (labeled as 1) or its absence (labeled as 0). Additionally, the severity of glioma is determined by the confidence levels associated with High-Grade Glioma (HGG), which includes grade III and IV gliomas associated with higher fatality rates, and Low-Grade Glioma (LGG), comprising grade I and II gliomas with generally longer patient survival. The images (in Figure 2) show the classification levels of low-grade glaucoma (LGG) and high-grade glaucoma (HGG), with each method highlighting significant features in the image representation.

In the image visualization, shown in Figure 2, the LGG level is greater than that of HGG, which shows that this region might not grow rapidly but is in a 3rd or 4th grade of glioma.

The image visualization shown in Figure 3 also says that the LGG level is greater than that of HGG, which shows that this region might not grow rapidly but is in a 3rd or 4th grade of glioma. Sample images for the LGG and HGG classifications are shown in Figure 4 and Figure 5, respectively.

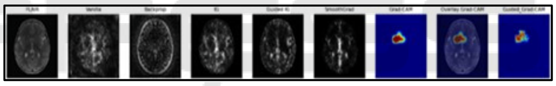

Figure 2: Predicted class 1, confidence of HGG: 0.3698, confidence of LGG: 0.63.

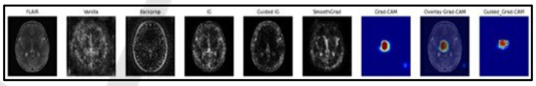

Figure 3: Predicted class 1, confidence of HGG: 0.4010, confidence of LGG: 0.59891.

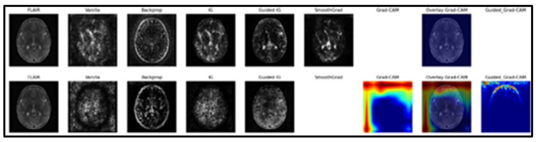

Figure 4: Samples images which are classified as LGG grade.

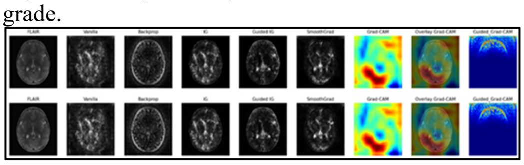

Figure 5: Sample images which are classified as HGG grade.

7.2 Segmentation based on UNet model

Here the deep brain glioma sub-region segmentation has been interpreted using multimodal MRIs from the BraTS 2021 validation dataset. The parameters are defined according to numerous factors, such as dimension, modality, XAI Mode, Class IDs, Tumor Label, Layer name, and segmentation model parameters, including dataset path and img_shape.

In the images shown in Figure 6, the XAI Mode specifies the mode of explainability. Here it is set to segmentation, as it deals with XAI methods applied to the segmentation process. The modality being used is FLAIR, a type of MRI sequence used in brain imaging that suppresses the effects of cerebrospinal fluid on the MRI image. The specific image ID for a particular MRI case is picked.

In the images shown in Figure 7, the XAI Mode is like the previous segmentation, but the modality is not specified to FLAIR alone, it may be either T1, TICE or T2. T1 shows the longitudinal relaxation, while TICE does the same with a contrast agent and finally T2 shows the transverse relaxation time.

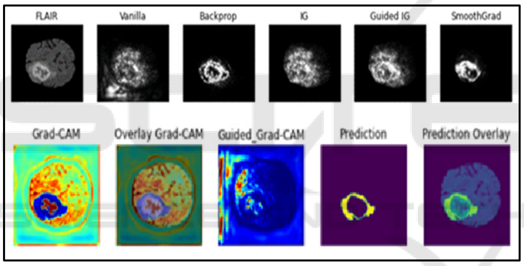

Figure 6: Example of segmentation with FLAIR modality and XAI methods applied.

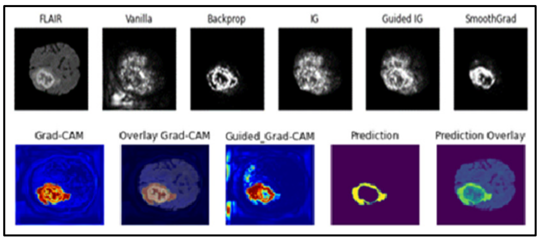

Figure 7: XAI Mode, Modality: All Models (FLAIR, T1, TICE, T2).

In the images of Figure 8, a set of IDs have been considered with 3 different MRI cases from the BraTS dataset. Here the last layer G-CAM has been highlighted shown in red regions which corresponds to a high score for the tumor region. Image has been iterated over the ID's 3 times for the set of tumor labels.

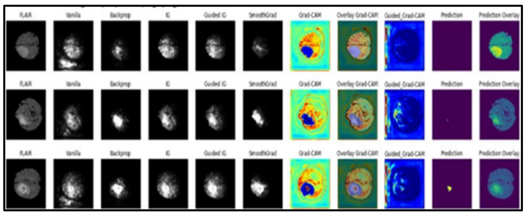

Figure 8: XAI IDs = ["BraTS2021 01652", "BraTS2021 00542", "BraTS2021 01381"], Modality: "FLAIR".

In the images of Figure 9, a set of IDs have been considered with 3 different MRI cases from the BraTS dataset classified as Good Slice IDs and the tumor labels are as a set of values 0,1,2,3. Here it considers all the modalities as dis-cussed earlier in 2nd case of images. The last layer, G-CAM, has been highlighted shown in red regions. This resulted high score for the regions with tumor. Image has been iterated over the IDs 3 times for the set of tumor labels.

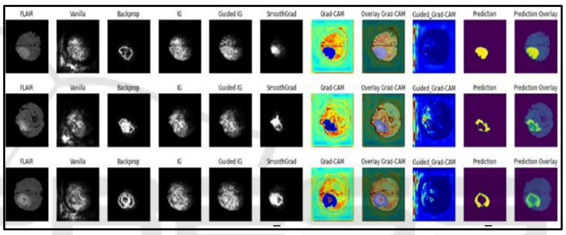

Figure 9: XAI IDs = ["BraTS2021 01652", "BraTS2021 00542", "BraTS2021 01381"], Modality: All Models.

7.3 Segmentation using CNN based UNet model

This segmentation result is done by providing information flow visualization on the internal layers of a segmentation CNN with the modality of FLAIR and the red regions comprising the XAI's last layer G- CAM. CNN follows a systematic approach for detecting brain gliomas by learning the abstract features available in the network such as the brain boundaries and identifies finely detailed tumor boundaries.

In the image visualization of Figure 10, a sample case has been considered from the BraTs dataset where segmentation have been done using CNN.

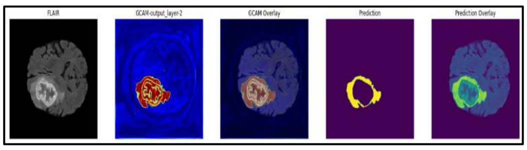

Figure 10: Segmentation CNN (Sample case).

With the IDs collected from the BraTS dataset, it is partitioned into classification and segmentation. From the classification method, the VGG19 model was processed comprising HGG and LGG and from the segmentation method the UNet model was processed comprising T1, TICE, T2 and FLAIR modalities.

8 CONCLUSIONS

The diagnosis of brain tumors from medical images is critical, as the images are varying greatly. The availability of convolution neural networks (CNNs) makes the brain tumor detection task easier. The CNN based approaches for brain tumor classification have paved the way for better tumor detection with increased accuracy. Using MRI images for detecting and classifying brain tumors is the recent focus. Interestingly, the combination of more than one types of CNN models has proved their performance for better feature extraction from the MRI images. In this study, a CNN is designed for brain tumor detection. The de- signed net-work is trained using two pretrained models - VGG19 and UNet, for faster and more convenient training. VGG19 is known for its deep architecture, while UNet is renowned for its capability in semantic segmentation tasks, making them a complementary and potent pair for brain tumor classification. This paper's aim is to detect brain tumor using this fusion model along with XAI methods to give proper visualization about how the tumor is detected with their own significant methods. The use of MRI as an imaging modality ensures detailed and informative data for accurate classification. Thus, this combinational approach addressed the concerns related to the "black box" nature of deep learning models in medical ap- plications.

In summary, the success of combining different convolutional models along with the methods of XAI suggests the potential for further exploration of fusion strategies in neural networks for medical imaging. This study contributes to the advancement of brain tumor detection methodologies, providing a foundation for future research.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to their institution (Amrita School of Computing, Amrita Vishwa Vidyapeetham, India), for the support provided in completing this project and paper submission.

REFERENCES

Abhilasha, K., S. Swati, and M. Kumar (2022). "Brain Tumor Classification Using Modified AlexNet Network". In: *Advances in Distributed Computing and Machine Learning*. Springer.

Amran, G.A. et al. (2022). "Brain Tumor Classification and Detection Using Hybrid Deep Tumor Network". In: *Electronics* 11.3457.

Apostolopoulos, I.D., S. Aznaouridis, and M. Tzani (2023). "An Attention-Based Deep Convolutional Neural Network for Brain Tumor and Disorder Classification and Grading in Magnetic Reso- nance Imaging". In: *Information* 14.174.

Buda, Mateusz (2022). *LGG MRI Segmentation Dataset*. https : / / www . kaggle . com / datasets / mateuszbuda/lgg-mri-segmentation.

Duvvuri, K. et al. (2022). "Grad-CAM for Visualizing Diabetic Retinopathy". In: *2022 3rd International Conference for Emerging Technology (IN- CET)*, pp. 1–4.

Hu, H. et al. (2021). "Brain Tumor Diagnose Applying CNN through MRI". In: *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*.

Maheswari, B.U. et al. (2023). "In-Hospital Mortality Prognosis: Unmasking Patterns Using Data Science and Explainable AI". In: *2023 9th International Conference on Signal Processing and Communication (ICSC)*, pp. 356–361.

Nair, Priyanka C. et al. (2023). "Building an Explain- able Diagnostic Classification Model for Brain Tumor Using Discharge Summaries". In: *Proce- dia Computer Science* 218, pp. 2058–2070.

Pathak, A., M. Kamani, and R. Priyanka (2023). "Brain Tumor Segmentation Using Modified Re- sUNET Architecture". In: *2023 International Conference on Sustainable Communication Networks and Application* (ICSCNA).

Priya, A. and V. Vasudevan (2024). "Brain Tumor Classification and Detection via Hybrid AlexNet- GRU Based on Deep Learning". In: *Biomedical Signal Processing and Control* 89.

Rani, N.S. et al. (2022). "Multi-Disease Diagnosis Model for Chest X-ray Images with Explain- able AI – Grad-CAM Feature Map Visualization". In: *2022 International Conference on Futuristic Technologies (INCOFT)*, pp. 1–5.

Schettler, David (2021). BraTS 2021 Task 1 Dataset. https://www.kaggle.com/datasets/dschettler8845/ brats-2021-task1.

Shah, Arya (2019). *Brain Tumor Segmentation - BraTS 2019*. https://www.kaggle.com/datasets/ aryashah2k / brain - tumor - segmentation - brats - 2019/data.

Tas¸cı, B. (2023). “Attention Deep Feature Extraction from Brain MRIs in Explainable Mode: DGX- AINet”. In: *Diagnostics (Basel)* 13.5, pp. 859.

Velden, Bas H.M. van der et al. (2022). “Explainable Artificial Intelligence (XAI) in Deep Learning- Based Medical Image Analysis”. In: *Medical Im- age Analysis* 79.

Younis, A. et al. (2022). “Brain Tumor Analysis Using Deep Learning and VGG-16 Ensembling Learning Approaches”. In: *Applied Sciences* 12.7282.

Zeineldin, R.A., M.E. Karar, Z. Elshaer, et al. (2022). “Explainability of Deep Neural Networks for MRI Analysis of Brain Tumors”. In: *International Journal of Computer Assisted Radiology and Surgery (CARS)* 17, pp. 1673–1683.