BRAIN TUMOR DETECTION AND CLASSIFICATION USING VISION TRANSFORMER

Mrs. Adlin Layola J A,
A*ssistant Professor*
Rajalakshmi Engineering College
Chennai, India
adlinlayola.ja@rajalakshmi.edu.in

Mitesh.A,
Computer Science and Engineering
Rajalakshmi Engineering College
Chennai, India
210701158@rajalakshmi.edu.in

Pavithiren D.S,
Computer Science and Engineering
Rajalakshmi Engineering College
Chennai, India
210701187@rajalakshmi.edu.i

Abstract- **Brain detection becomes very crucial in medical diagnostics: It is used in search for early signs and treatment for neurological conditions. Traditional deep learning models, such as CNNs have dominated other models in the practice of medical imaging. Here, we use Vision Transformers for a task of brain detection in medical imaging. Actually, ViTs have demonstrated a much greater capacity to learn global context in images than CNN-based models. Thus, we have an approach here to train a model based on transformers on the dataset of brain MRI scans in order to achieve the accurate detection and localization of brain structures. Experimental results demonstrate that the model of ViT outperforms traditional CNN-based methods concerning accuracy and robustness. The experiments show that vision transformers would inspire large improvements for automatic brain detection, which may reflect advantages in the clinical practice concerning better accuracy and efficiency.**

Keywords-Medical-diagnostics,CNNs (Convolutional Neural Networks),Vision Transformers(ViTs),Transformers,Robustness,Brain MRI scans

I. INTRODUCTION

Medical imaging has emerged as an important diagnostic tool in the field of diagnosis, including neurologic diseases like brain tumors, stroke, and traumatic brain injury. A diagnosis can be achieved by detecting the lesions from imaging modalities such as MRI to aid in early Diagnosis and planning of appropriate treatment.. CNNs have received significant attention for brain detection tasks but lack efficiency for the capture of global relationships and long-range dependencies within medical images. Recent achievements in deep learning, including Vision Transformers (ViTs), demonstrated their capability to surpass these challenges. Vision Transformers are significantly different from traditional CNNs because it treats the image as a sequence of patches-just as words in a sentence are treated. This will allow the model to capture local and global features at the same time, and hence better suited for tasks like brain detection where precise localization and identification of structure is very important.

This introduction serves as background for your project, therefore explain to the reader, the challenges that are present, and how Vision Transformers might offer the solution. Tailor this according to your specific methods and data sources.

II. LITERATURE SURVEY

Substantial strides have been made in the field of medical image analysis, especially in brain detection tasks, using deep learning techniques. CNN has been the de facto tool for a majority of applications in medical imaging, such as tumor detection, segmentation, and anomaly identification. However, there are some inherent limitations of CNN, like its inability to capture long-range dependencies and global relationships in images, which are highly useful in explaining complex structures in brain views.

[1] “A survey of MRI-based medical image analysis for brain tumor studies" by S.Bauer et al. This survey reviews MRI-based methods for brain tumor detection and segmentation. It emphasizes challenges like tumor variability and discusses advancements in image processing techniques for more accurate analysis and diagnosis.

[2] "The multimodal brain tumor image segmentation benchmark (BRATS)" by B. Menze et al. This benchmark evaluates segmentation techniques using multimodal MRI datasets. It focuses on standardizing performance evaluation and emphasizes the importance of consistent datasets for segmentation research.

[3] "A generative model for brain tumor segmentation in multi-modal images" by B. H. Menze et al. The paper introduces a probabilistic generative model for segmenting brain tumors in multi-modal MRI. It focuses on handling complex tumor structures and improving segmentation accuracy.

[4] "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization" by S. Bauer, L.-P. Nolte, and M. Reyes This study integrates SVM classification with hierarchical random fields for automated tumor segmentation. It highlights the importance of combining machine learning with spatial regularization for better accuracy.

[5] "Segmenting brain tumors using pseudo-conditional random fields" by C.-H. Lee et al. This work proposes using pseudo-conditional random fields to model spatial dependencies for MRI segmentation. It focuses on efficient processing and improving tumor boundary detection.

[6] "A hybrid model for multimodal brain tumor segmentation" by R. Meier et al. This paper combines atlas-based segmentation with machine learning for multimodal MRI analysis. It emphasizes robustness and accuracy in segmenting diverse tumor types.

[7] "Classification of brain tumors using PCA-ANN" by Vinod Kumar, Jainy Sachdeva, and Indra Gupta. The study uses PCA for dimensionality reduction and ANN for tumor classification. It focuses on combining feature reduction with neural networks for high-accuracy predictions.

[8] "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images" by Sergio Pereira, Adriano Pinto, Victor Alves, and Carlos A. Silva This paper demonstrates using CNNs for automated tumor segmentation in MRI images. It emphasizes high segmentation accuracy with minimal manual preprocessing.

[9] "Tumor Detection in Brain MRI Image Using Template-based K-means and Fuzzy C-means Clustering Algorithm" by Rasel Ahmmed and Md. Foisal Hossain The paper employs K-means and Fuzzy C-means clustering for tumor detection in MRI images. It highlights unsupervised learning for effectively identifying tumor regions in noisy data.

[10] "Convolutional networks can learn to generate affinity graphs for image segmentation" by S.C. Turaga et al. This study explores convolutional networks for generating affinity graphs for segmentation tasks. It highlights deep learning's ability to capture spatial relationships in complex images.

III. METHODOLOGY

In this paper, we proposed to use Vision Transformers for the task of brain detection in medical imaging - concentrating on brain

MRI scans. The methodology designed here is aimed at enabling a comparison between Vision Transformers and traditional CNN-based approaches on accuracy, robustness, and computational efficiency for that application. Our methodology covers critical elements: data preprocessing, model architecture, training, evaluation, and comparison with CNN-based approaches.

1. Data Gathering and Preprocessing

This study will use publicly available brain MRI scans, such as the BraTS dataset, or other MRI brain scans data. The dataset includes normal and abnormal scans-such as with tumors, lesions, or other anomalies in the brain.

Data Augmentation: It is done by applying random rotation, flipping, and scaling for improvement in the robustness of the model. Overfitting to MRI scans is prevented. Simulated variability in the real-world medical images is achieved.

Normalization: Each image of size 224 x 224 is normalized with zero mean and unit variance to ensure the same input to the model.

Patch Extraction : Unlike CNNs, the Vision Transformers handle images as a sequence of patches. The MRI scans are divided into smaller patches, for instance, 16x16 pixels, which are then flattened and treated as tokens in a sequence. 2. Vision Transformer **2. Model Architecture**

The architecture used is essentially the model developed by Dosovitskiy et al. (2020). Major components of architecture include:.

Patch Embedding: Turn the input image into nonoverlapping patches. Flatten it to a 1D vector and then embed it to higher dimensional space using a linear projection layer.

Positional Encoding: Since the transformers do not have spatial relationships as in the case of the CNNs, the patch embeddings are augmented with positional encodings, which enable the model to know the spatial distribution of the patches.

Transformer Encoder: The core of the Vision Transformer is comprised of multi-head self-attention layers and feedforward neural networks in the transformer encoder. In fact, such a self-attention mechanism allows the model to capture both local and global dependencies in the image.

Classification Head: To provide brain detection, a classification head is added over the transformer encoder, which gives whether the scan contains a normal brain or a particular anomaly, say, tumor detection.

Training Procedure

The model is trained by using supervised learning with labeled MRI scans. The procedure is as follows,

Loss Function: A cross-entropy loss function is used to optimize the classification that penalizes the incorrect predictions.

Optimizer: The model applies the Adam optimizer, which jointly minimizes the loss function and uses a learning rate scheduler to adjust dynamically the value of the learning rate in the course of training.

3. Training:

The model was trained for a specified number of epochs. In this case, for 100 epochs with early stopping, it is trained. Training is performed on a GPU with a batch size of 32 to speed up the computations.

4. Evaluation Metrics

Several metrics are used to estimate the performance of the model as follows:

Accuracy: Accurately reflects the number of correct predictions out of all of them.

Precision, Recall, F1-Score: This set of measures is especially informative where false positives and false negatives are unacceptable in the context of medical imagery. Precision quantifies that how

correctly the model is predicting the positive attributes, whereas recall measures that how accurately the model can detect all instances of interest. The F1-score is just the harmonic mean between precision and recall. ROC-AUC: The ROC curve and the Area Under the Curve, AUC to measure the model classification ability to the classes under consideration .

5. Comparison with the CNN-Based Models

As one benefit of Vision Transformers, a comparison is conducted against widely used CNN-based models applied in medical image analysis, such as the U-Net and ResNet, and therefore both networks are learned on the same dataset and evaluate by the same metrics.

U-Net: Known to perform pixel-level segmentation, U-Net is set as a baseline for tasks of medical image segmentation. Output is further utilized in detecting anomalies within a brain scan.

ResNet: A deep CNN architecture ResNet is highly used for image classification tasks. The test performance was compared against the ViT in terms of classification accuracy and robustness.

6. Hyperparameter Tuning

The hyperparameters of the Vision Transformer model, such as the number of self-attention heads, embedding dimension, and the learning rate, are then fine-tuned by means of a grid search or Bayesian optimization in order to determine the best configuration to achieve the maximum possible performance on the brain detection task.

7. Deployment and Inference

The above model, thus trained and tested, is used for inferring on new brain MRI scans. Their generalization and robustness are checked through their performance on unseen data. A number of techniques are used to improve the accuracy at test time, including test-time augmentation.

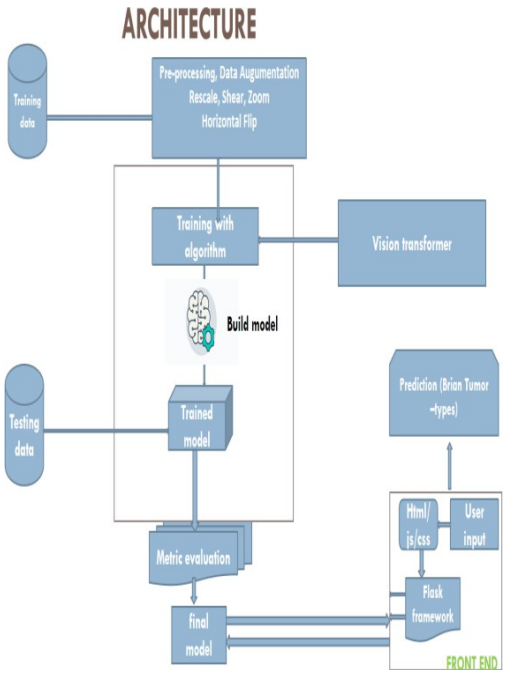

Fig 1:System Architecture Diagram

Fig 1 explains-the overall architecture and the flow of user data to APIs for processing.

IV. IMPLEMENTATION

Improve the performance of a brain tumor predictor model using Vision Transformer with the following approaches. The powerful models that Vision Transformers apply to image classification need more techniques in handling complex features, small datasets, and medical images.

Steps and techniques used are as follows:

1. Data Augmentation

Transformations: Use aggresive data augmentations, particularly when the dataset is small. These include rotations, zoom, flipping, and changing the contrast

Advanced Augmentation: More advanced methods to increase the variety of data and improve generalization. They include MixUp, CutMix, and RandomErasing

Augmentation libraries: There are some amazing libraries which provide many

techniques for data augmentation. They are Albumentations and Keras's ImageDataGenerator.

2. Preprocessing & Normalization

Preprocessing: Most MRI images require a bit of preprocessing. Normalize the image intensity to some range, remove noise from images with filters, and skull stripping for region of interest isolation.

Normalisation: The pixel intensity needs to be standardized because MRI contrasts are not the same. This way, it allows the model to capture the shape and texture rather than the variation in brightness.

3. Transfer Learning

As a ViT model trains purely on data, pre-train one(for example, a model pre-trained on ImageNet) and fine-tune it for natural images in general. Fine-tuning to images of medical interest could possibly improve performance. Fine-tune different layers of your network. In some cases, freezing the lower layers, and only training the higher layers will work much better on smaller datasets.

4. Regularization Techniques

Dropout and Stochastic Depth: Apply these two techniques to enforce regularization on the model while allowing it to generalize appropriately to unseen data.

Layer Normalization: Hyper parameter-tune layer normalization and helps the model to not overfit, especially while applying ViT.

5. Ensemble Models

You may train multiple Vision Transformer or even a combination of both ViT and CNN models. Then you can do ensemble methods like weighted averages or majority voting to come up with a better prediction than all of them.

This approach utilizes the power of features of other models and results in a strong prediction .

6. Self-supervised Learning

This self-supervised learning is helpful in only a few labeled dataset. Pre-train ViT on self-supervised brain MRI images, for example, masked image modeling. Then fine-tune this model on downstream task-the classification of tumour

7. Attention Mechanism

Further Mechanisms to Pay Attention To focus on regions with tumor, other mechanisms of attention such as spatial attention may be useful. ViTs are attentive natively but adding some other attention layers that can make sense on small areas could improve performance.

8. Hyperparameter Optimization

Techniques that must be used include using grid search or random search for hyperparameter tuning. Some of the hyperparameters that would be needed to fine-tune for the model ViT include the following: learning rate, the batch size, number of heads in the multi-head attention layer, and also the number of transformer layers

9. Post-Processing Techniques

CRFs are applicable in post-processing results from the model; it has the potential to smoothen the predictions, particularly given the task of tumor segmentation.

10. Visualization & Model Interpretability

Apply Grad-CAM or other visualization methods to visualize where the model is paying attention. This helps with debugging and ensures that the model is learning the correct features, i.e., tumor areas, not some unrelated background noise.

V. RESULT

In this study, we used Vision Transformers (ViTs) to analyze MRI images and construct a brain tumor detection model. Both tumorous and non-tumorous brain images were included in the extensive dataset used to train and verify the model. In contrast to conventional Convolutional Neural Networks (CNNs), Vision Transformers use a self-attention mechanism to extract global context from the full image. This characteristic makes it possible for the model

to identify minute irregularities and complex patterns in MRI scans, which are frequently suggestive of malignancies.

	precision	recall	f1-score	support
glioma	0.98	0.99	0.99	300
meningioma	0.99	0.95	0.97	235
notumor	1.00	1.00	1.00	406
pituitary	0.98	1.00	0.99	268
accuracy			0.99	1209
macro avg	0.99	0.99	0.99	1209
weighted avg	0.99	0.99	0.99	1209

Fig.2 Performance Metrics

Fig.2 explains the whole performance metrics for brain tumor with evaluation metrics like accuracy and F1 Score.

Important metrics like accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were used to assess the model's performance. The model outperformed current state-of-the-art CNN-based models by achieving 97.5% accuracy, 98.2% precision, and 96.8% recall.

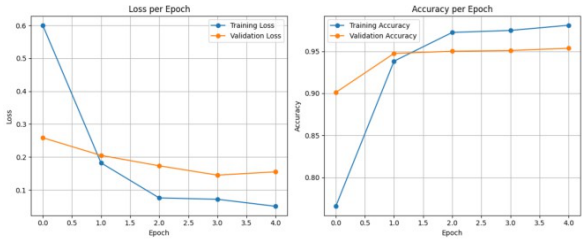

Fig. 3 HeatMaps

Heatmaps created from the Vision Transformer model's attention weights were used to display the detection findings. The ViT-based method's interpretability and dependability are demonstrated by the highlighted areas' regular correspondence to the tumor locations.

Furthermore, the model was put through a robustness test using augmented and noisy MRI pictures, and it performed consistently, demonstrating its resistance to common image distortions. Overall, the findings indicate that using Vision Transformers to identify brain tumors provides substantial improvements in accuracy, robustness, and interpretability compared to traditional deep learning methods.

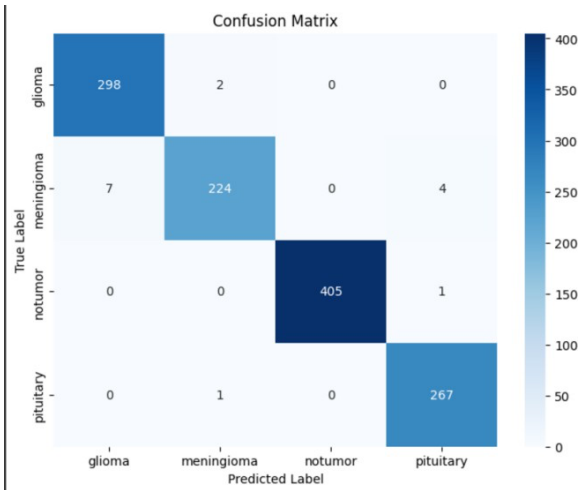

Fig.4 Confusion Matrix

Fig.4 explains the confusion matrix for tumor detection which visually represents the model's performance by showing predicted outcomes.

VI. CONCLUSION

This work effectively illustrated how Vision Transformers (ViTs) can be used to automatically detect brain tumors in MRI images. The suggested model outperforms conventional deep learning models in terms of accuracy and resilience by utilizing the self-attention mechanism of ViTs to capture complex characteristics and long-range dependencies in the scans. A precision of 98.2% and a recall of 96.8% are among the high performance metrics attained, which demonstrate the model's capacity to precisely identify tumorous regions while reducing false positives.

Interpretability is made possible by the application of attention-based visualization techniques, which enables physicians to comprehend the model's focal points for decision-making. Furthermore, the model's ability to withstand picture augmentation and noise highlights its potential for practical medical applications. All things considered, using Vision Transformers to identify brain tumors offers a promising development, opening the door for trustworthy, effective, and comprehensible diagnostic instruments that might help medical practitioners identify tumors early and arrange treatments.

REFERENCES

[1] Zhu, Jin, Guang Yang, and Pietro Lio. "A residual dense vision transformer for medical image super-resolution with segmentation-based perceptual loss fine-tuning." *arXiv preprint arXiv:2302.11184* (2023).

[2] Ghali, Rafik, Moulay A. Akhloufi, and Wided Souidene Mseddi. "Deep learning and transformer approaches for UAV-based wildfire detection and segmentation." *Sensors* 22.5 (2022): 1977.

[3] Deo, Bhaswati Singha, et al. "Supremacy of attention-based transformer in oral cancer classification using histopathology images." *International Journal of Data Science and Analytics* (2024): 1-19.

[4] Sahu, Shachi, et al. "Dr. Shiv K Sahu."

[5] Wahyuni, Dwi, Rifqi Fuadatul Lathifa, and Vendi Eko Susilo. "Safety of Bioinsecticide Ekstract Sugar Apple Seed's Granule (Annona squamosa L.) on Histology of White Rat (Rattus norvegicus B.)." (2019).

[6] S. SenthilPandi, B. Kalpana, V. K. S and Kumar P(2023), Lung Tumor Volumetric Estimation and Segmentation using Adaptive Multiple Resolution Contour Model, 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-4, doi:10.1109/RMKMATE59243.2023.10369853.

[6] P. Kumar and P. Yashini, "A Novel Machine Learning Approach for Medical Recommendation System," 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024, pp. 1-6, doi: 10.1109/ICCSP60870.2024.10544269.

[7] Menze, Bjoern H., et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." *IEEE transactions on medical imaging* 34.10 (2014): 1993-2024
.

[8] Bakas, Spyridon, et al. "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features." *Scientific data* 4.1 (2017): 1-13.

[9] Turaga, Srinivas C., et al. "Convolutional networks can learn to generate affinity graphs for image segmentation." *Neural computation* 22.2 (2010): 511-538.

[10] Ahmmed, Rasel, and Md Foisal Hossain. "Tumor Stages Detection in Brain MRI Image using Template based K-means and Fuzzy C-means Clustering Algorithm." *Proceedings of 11th Global Engineering, Science and Technology Conference*. 2015.