NEUROSCAN: Revolutionizing Brain Tumor Detection Using Vision-Transformer

Kamran Khan, Najam Aziz, Afaq Ahmad, Munib-ur-Rehman, Yasir Saleem Afridi
Department of Computer Systems Engineering, University of Engineering & Technology, Peshawar, Pakistan
*** Correspondence:** Kamran Khan (20pwcse1895@uetpeshawar.edu.pk)

Citation | Khan. K, Aziz. N, Ahmad. A, Rehman. M, Afridi. Y. S, "NEUROSCAN: Revolutionizing Brain Tumor Detection Using Vision-Transformer", IJIST, Special Issue pp 143-151, May 2024
Received | May 14, 2024, **Revised |** May 18, 2024, **Accepted |** May 20, 2024, **Published |** May 23, 2024.

Brain tumor detection is a pivotal component of neuroimaging, with significant implications for clinical diagnosis and patient care. In this study, we introduce an innovative deep-learning approach that leverages the cutting-edge Vision Transformer model, renowned for its ability to capture complex patterns and dependencies in images. Our dataset, consisting of 3000 images evenly split between tumor and non-tumor classes, serves as the foundation for our methodology. Employing Vision Transformer architecture, we processed high-resolution brain scans through patching and self-attention mechanisms. The model is trained through supervised learning to perform binary classification tasks. Our employed model achieved a high of 98.37% in tumor detection. While interpretability analysis was not explicitly performed, the inherent use of attention mechanisms in the Vision Transformer model suggests a focus on important brain regions and enhances its potential for prioritizing crucial information in brain tumor detection.

Keywords: Brain Tumor Detection, Medical Imaging, Classification, Vision Transformers, ViT, Machine Learning, Deep Learning.

Introduction:

Brain tumors are a major global health concern, and by early identification, patient healthcare can be greatly enhanced. Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are commonly used medical imaging techniques for tumor diagnosis in the human brain. However, the diagnosis still relies on the doctor or radiologist's decision, which is an error-prone and time-consuming process. In order to support medical personnel in making fast and precise diagnoses, the increasing amount of brain imaging data needs sophisticated computational algorithms. This research presents the incorporation of a type of artificial neural network called vision transformer (ViT) for brain tumor detection. ViT is a viable option for medical image analysis since they have demonstrated exceptional performance in a variety of computer vision tasks. The purpose of the research is to assess the accuracy of vision transformers while detecting brain tumors.

Vision transformers are a useful tool for detecting brain tumors because of their capacity to extract contextual information and spatial correlations from medical images. In contrast to conventional convolutional neural networks (CNNs), which analyze images using a fixed grid-like structure, vision transformers use a self-attention mechanism that enables them to take interactions and long-range relationships between various image regions into account. Additionally, vision transformers have proven to perform remarkably well in a variety of computer vision applications, such as segmentation, object identification, and image classification. They are a desirable alternative for medical image analysis, where reliable and accurate abnormality detection is crucial, due to their capacity to learn from large-scale datasets and generalize well to unseen data.

This research aims to assess the effectiveness of ViT for identifying brain tumors through extensive experimentation and achieve an acceptable accuracy by ViT for the said task. Our research will demonstrate the potential of ViT as a valuable tool for diagnosing brain tumors in the clinical setting. The research paper has been divided into five sections. Section 1 elaborates on the Introduction, Section 2 covers the Literature Review, Section 3 covers the methodology, Section 4 discusses results and Section 5 concludes the paper.

Literature Review:

During the previous decade, deep learning algorithms made great hype for their high performance in various domains ranging from textual analysis to image analysis. These algorithms are also extensively used for healthcare data and help in automating various tasks related to healthcare while achieving comparable accuracy to that of humans. The incorporation of deep learning techniques has resulted in a remarkable revolution in the field of brain tumor identification. Conventional approaches in this field mostly depend on manually designed features and rule-based algorithms, which frequently prove inadequate for managing the complex and diverse nature of medical images. In a research study, Kumar et al. [1], have made significant progress in proving that Convolutional Neural Networks (CNNs) are effective at automating the process of identifying brain tumors. Conventional methods have difficulties when it comes to capturing spatial hierarchies and patterns inside medical images. CNNs have proven essential in overcoming these challenges.

Even with CNN's significant success, it is still necessary to investigate new approaches. Hossain et al. [2] developed the concept of vision transformers and provided a new dimension to improve the detection of brain tumors. Vision transformers provide a comprehensive viewpoint by removing global dependencies from images, which may enhance the model's capacity to identify complex and subtle patterns suggestive of malignancies. By customizing vision transformers for the classification of medical images, Manzari et al. [3] have made a significant contribution to this investigation. Their research highlights how adaptable these transformer models are at handling the difficulties presented by intricate medical imagery.

The quick advancement of medical imaging technology emphasizes how crucial it is to identify brain tumors with accuracy. Puttagunta et al. [4] offer an extensive overview of deep learning methods for medical image interpretation. Their work describes the changing environment in which precise identification becomes more and more important, in addition to highlighting the revolutionary potential of these tools. Furthermore, the work of Sadad et al. [5] offers insightful information about brain tumor detection through a wider variety of machine learning algorithms, offering a comprehensive overview of the difficulties faced and prospects in the field.

Ahn et al. [6] in their study, provide the theoretical foundation for comprehending attention mechanisms, which are a crucial aspect of vision transformers. The model can detect global dependencies in images, thanks to this attention mechanism, which is in line with the complex needs of brain tumor diagnosis. The field of brain tumor diagnosis has experienced a dynamic shift due to the incorporation of state-of-the-art deep learning techniques. This development was prompted by the limits of traditional approaches that relied on manually created features and rule-based algorithms. Recent research has demonstrated the revolutionary potential of convolutional neural networks (CNNs) in automating the complex process of recognizing brain tumors, as demonstrated by the work of Kumar et al. [1]. With their capacity to extract intricate patterns and spatial hierarchies from medical images, CNNs are a key component in the development of diagnostic techniques.

The study of Brownlee et al [7], focuses on mastering machine learning techniques by providing useful information that can support the continued advancement and deployment of deep learning techniques in the particular field of brain tumor identification. The notion of vision transformers was created by Akinyelu et al. [8] added a new layer to enhance ViT for brain tumor identification. By eliminating global dependencies from images, vision transformers offer a holistic perspective that may improve the model's ability to recognize intricate and nuanced patterns suggestive of tumors. A noteworthy contribution to this study has been made by Wanget al. [9], who customized vision transformers for the classification of medical images. Their study demonstrates how versatile these transformer models are in managing the challenges posed by complex medical images.

Jiang, et al. [10], in their work, provided an in-depth review of deep learning techniques for the interpretation of medical imaging data. Their work highlights the revolutionary potential of these techniques and outlines the evolving environment in which exact identification becomes increasingly vital. Moreover, Tummala et al. [11] provide informative data regarding the identification of brain tumors using a greater range of machine learning algorithms, providing a thorough summary of the challenges encountered and future prospects in the field. Asiriet al. [12] utilized an attention mechanism to enable the model to identify global dependencies in images, which is consistent with the requirements of the complicated task of brain tumor diagnosis. The work of [13][14][15][16] provides a variety of perspectives and approaches that greatly enhance our knowledge and progress in the field of brain tumor detection using deep learning techniques.

Methodology:

This section explains the overall methodology of this research and is presented in Figure 1 by the methodology flow diagram.

Data Description:

The dataset that has been utilized for our research study is the Br35H Brain Tumor Detection 2020 Dataset which is available on Kaggle. The dataset contains 3000 T1 weighted images of tumorous (yes) class and healthy (no) brain scans. Each class has 1500 images and an equal distribution of the Dataset. Sample MRI brain scans of tumors and without tumors are shown in Figures 2(a) and (b), respectively.

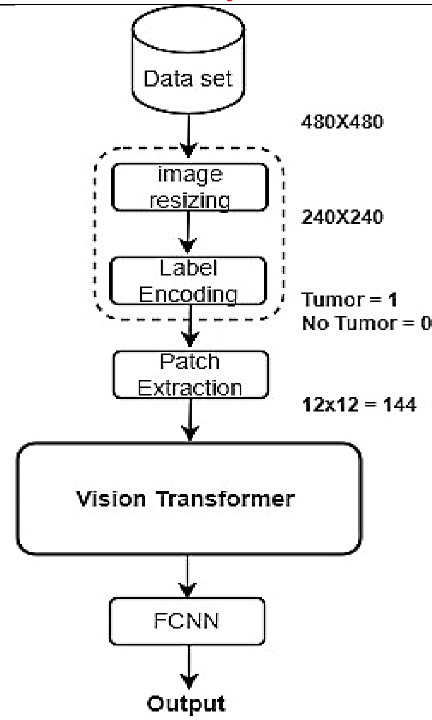
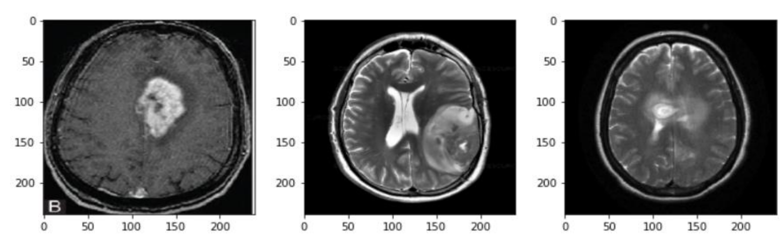
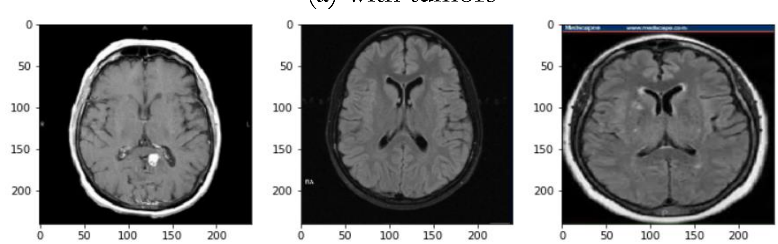


Figure 1: Methodology flow diagram



(a) with tumors



(b) without tumors

Figure 2: Samples from the dataset

Preprocessing:

Data preprocessing is a vital step in preparing data that is suitable for model training. This enhances data quality, and compatibility and ensures optimal performance by the model being trained. Various preprocessing steps that are applied to the utilized dataset are explained as follows.

Resizing:

The dataset contains images that are of varying sizes. Hence, resizing is done to bring every image to 240 × 240 pixels, which was an important first step. This consistency allowed our model to have consistent input dimensions, which made processing easier.

Normalization:

To scale data to a common range, normalization is done. The process ensures that all the input features have a similar influence on the model, preventing certain features from dominating the learning process due to differences in their original scales. This aids in effective learning and performance improvement of the model. To achieve normalization, we used the following mathematical transformation.

$$X' = \frac{X - Xmin}{Xmax - Xmin}$$

Where the resized image is represented by X and the normalized image is represented by X′.

Label Encoding:

By label encoding, we converted the categorical labels yes and no for brain tumor presence and absence, into numerical values to make sure they are compatible with our model.

Vision Transformer (ViT):

The Vision Transformer, or ViT, is a model for image classification that employs a Transformer-like architecture over patches of the image. An image is split into fixed-size patches, each of them is then linearly embedded, position embeddings are added, and the resulting sequence of vectors is fed to a standard Transformer encoder. In order to perform classification, the standard approach of adding an extra learnable "classification token" to the sequence is used [17]. The intricate details of medical images are catered to in the vision transformer architecture. In order to process image patches and enable the network to capture both local and global information, the model is composed of a number of transformers blocks as shown in Figure 3. The model can assess the relative importance of various regions in the input images using the self-attention process.

Model Architecture:

The model architecture we have implemented for this study is depicted in Figure 3. and is explained as follows.

- **Image Patching**: The input to the model is split up into smaller patches as part of the patching process. This separation makes processing more efficient and makes it possible for the model to successfully collect local spatial information.
- **Image Flattening:** After patching, the two-dimensional spatial information is converted into a linear representation by flattening the patches. The data is now ready for additional processing inside the model.
- **Patch Encoder:** The Patch Encoder module processes the flattened patches. This module's dense projection layer and positional encoding encode the patches' spatial information and give the locations of the patches within the image context.
- **Positional Encoding:** An essential part of embedding positional information into the patch representations is positional encoding. As a result, the model is able to comprehend the spatial relationships between patches and gather crucial categorization context.
- **Transformer Encoder Block:** The Transformer Encoder Blocks are the central component of the model. These blocks are made up of feedforward neural network layers and multi-head self-attention layers. They facilitate the model's ability to extract high-level characteristics from the encoded patches and to capture global dependencies.
- **Aggregate Representation:** Upon completion of several Transformer Encoder Blocks, the representations obtained from every patch are combined. The contextual data that is acquired from every patch is combined throughout this aggregation phase to provide a complete representation of the full image.
- **Classification:** In the end, a classification head made up of dense layers receives the combined representation. To ascertain whether a brain tumor is present in the input image, these layers carry out categorization.

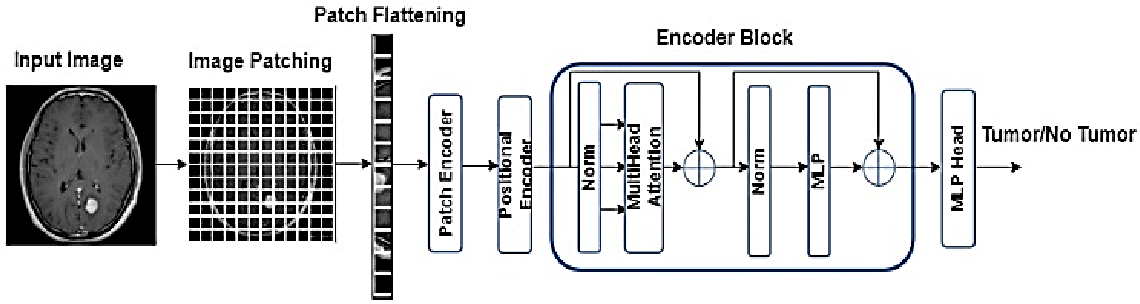


Figure 3: Model Architecture

Training Strategy:

Here, we describe the training approach that we used to train our brain tumor identification model using the Br35H: Brain Tumor Detection 2020 Kaggle dataset in an efficient manner.

Data Splitting:

We divided the dataset into separate training and testing sections in order to create a strong training regimen. To ensure there was enough data for the model to learn, we set aside 80% of the dataset for training. The remaining 20% was set aside for testing, allowing for an objective assessment of the performance of the trained model.

Hyperparameters Settings:

- **Learning Rate:** We regulate the rate at which the model modifies its parameters during optimization by setting the learning rate to 0.001. The trade-off between parameter stability and convergence speed is balanced by this decision.
- **Weight Decay:** To prevent overfitting inclinations, a weight decay of 0.0001 was used as a regularization strategy. Model generalization is promoted by weight decay, which penalizes large parameter values.
- **Batch Size:** The number of samples processed in each training iteration was determined by our training procedure, which included a batch size of 32. This batch size balances gradient stability with computational efficiency.
- **Number of Epochs:** The training process took place over fifty epochs, allowing for iterative dataset learning. The model is exposed to the data throughout several epochs, which promotes model convergence and parameter refining.
- **Image Size and Patch Size**: After resizing the input images to 240 by 240 pixels, patches with a size of 20 by 20 pixels were extracted. These dimensions played a crucial role in specifying the input organization and spatial resolution of the feature extraction process of the model.
- **Transformer Architecture:** Eight transformer layers, each incorporating four attention heads, made up our model architecture. Dense layers with units [2048, 1024] were featured in the final classifier, while the transformer units had size [128, 64]. These architectural decisions were carefully considered in order to maximize the expressiveness of the model and its ability to identify complex patterns in the data.

Result and Discussion:

The implemented ViT model shows impressive performance while utilizing the Br35H dataset. The model performance plots are shown in Figure 4. and the classification report is shown in Table 1.

Training and Validation Loss Graph:

The Graph shows how training and validation loss changed over the course of 50 epochs while our brain tumor detection model was being trained. At first, there is a declining trend in both the training and validation losses, which suggests that the model is successfully learning from the data. Both trajectories, however, converge at the 50th epoch, indicating that the model has reached a stage at which more training no longer considerably enhances its performance. This convergence shows that the model is performing at its best and has successfully captured the underlying patterns in the data. The losses then level out or maybe even begin to rise, indicating that the model has taken in all the information it can from the training set. Overall, the convergence after 50 epochs shows that training was successful and efficient by the model.

Training and Validation Accuracy Graph:

The accuracy of our brain tumor detection model during training and validation is shown in the graph. Both accuracy levels rise gradually, peaking close to the 50th epoch. At this stage, the model obtains an impressive testing accuracy of 98.37%, demonstrating its efficacy in correctly categorizing photos of brain tumors. This great accuracy shows how reliable and appropriate the model is for practical use in clinical settings.

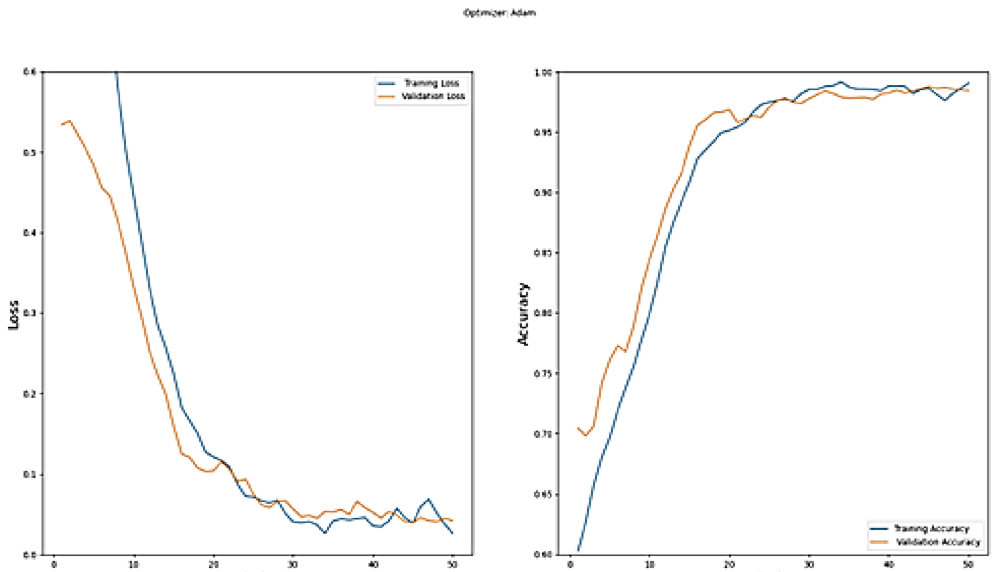


Figure 4: ViT Training and validation (a) loss (b) accuracy

Table 1: Classification Report

	Precision	Recall	F1-Score	Support
Yes	1.00	0.99	0.99	300
No	0.97	0.99	0.98	300
Accuracy			0.98	600

Explanation of Classification Report:

- **Precision:** Precision is defined as the percentage of actual positive forecasts among all positive predictions. The precision for the "Yes" class (brain tumor presence) in this instance is 1.00, meaning that all of the cases that were predicted to have brain tumors were accurate. Similarly, the precision for the "No" class (absence of brain tumor) is 0.97, meaning that 97% of cases that were predicted to be brain tumor-free were accurate.
- **Recall:** The percentage of true positive predictions among all actual positive cases is measured by recall, which is sometimes referred to as sensitivity. With a recall of 0.99 for the "Yes" class, the model was able to accurately identify 99% of real brain tumor cases. The recall for the "No" class is likewise 0.99, meaning that 99% of real cases.
- **Support:** In the dataset, Support is the total number of instances of each class. Our test data consists of 300 MRI scans from each class, totaling 600.
- **Accuracy:** Measured as the percentage of correctly classified occurrences over the total number of examples, accuracy assesses the overall correctness of the model's

predictions. With an accuracy of 0.98, 98% of the cases in the dataset were properly classified by the model.

Conclusion:

The field of medical image processing is undergoing a fundamental paradigm change with the integration of vision transformers for brain tumor diagnosis. Vision transformers are extremely useful tools for healthcare practitioners because of their special qualities, which include their interpretability and ability to understand complex relationships among images. One important function of the attention mechanism, which is a distinguishing characteristic of vision transformers, is to shed light on the characteristics that influence the model's judgment. In the medical industry, confidence and comprehension of the diagnostic procedure are greatly enhanced by transparency. The vision transformer model is being improved and optimized on a continuous basis. To evaluate the model's generalizability and robustness, extensive research is being conducted across a wide range of datasets and clinical settings.

Finally, this study emphasizes how important it is to incorporate vision transformers into the field of brain tumor identification. The model's proven performance on a variety of complex medical images represents a significant breakthrough in diagnostic skills. In addition to their exceptional performance, vision transformers can be interpreted, giving medical professionals useful information on the decision-making process. In addition to having the capacity to interpret intricate patterns, these models have the potential to improve accuracy, which makes them revolutionary instruments for medical practitioners. This study represents not only a significant advancement but also a possible paradigm changes in the field of brain tumor detection.

Acknowledgement:

Our project's success is mainly due to the outstanding advice and assistance of several people. We sincerely thank our Supervisor, Dr Yasir Saleem Afridi, and our Advisor, Engineer Najam Aziz, for their invaluable support. We also acknowledge the University of Engineering and Technology Peshawar, our parents, and our instructors for their constant support and for creating a supportive environment in which we may work.

Author's Contribution:

All the Authors have contributed equally to this work.

Conflict of Interest: The authors hold no conflict of interest in publishing this manuscript in IJIST.

References:

[1] S. Kumar, R. Dhir, and N. Chaurasia, "Brain Tumor Detection Analysis Using CNN: A Review," Proc. - Int. Conf. Artif. Intell. Smart Syst. ICAIS 2021, pp. 1061–1067, Mar. 2021, doi: 10.1109/ICAIS50930.2021.9395920.

[2] S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab, and M. J. Piran, "Vision Transformers, Ensemble Model, and Transfer Learning Leveraging Explainable AI for Brain Tumor Detection and Classification," IEEE J. Biomed. Heal. Informatics, vol. 28, no. 3, pp. 1261–1272, Mar. 2024, doi: 10.1109/JBHI.2023.3266614.

[3] O. N. Manzari, H. Ahmadabadi, H. Kashiani, S. B. Shokouhi, and A. Ayatollahi, "MedViT: A robust vision transformer for generalized medical image classification," Comput. Biol. Med., vol. 157, p. 106791, May 2023, doi: 10.1016/J.COMPBIOMED.2023.106791.

[4] M. Puttagunta and S. Ravi, "Medical image analysis based on deep learning approach," Multimed. Tools Appl. 2021 8016, vol. 80, no. 16, pp. 24365–24398, Apr. 2021, doi: 10.1007/S11042-021-10707-4.

[5] T. Sadad et al., "Brain tumor detection and multi-classification using advanced deep learning techniques," Microsc. Res. Tech., vol. 84, no. 6, pp. 1296–1308, Jun. 2021, doi: 10.1002/JEMT.23688.

[6] K. Ahn, X. Cheng, M. Song, C. Yun, A. Jadbabaie, and S. Sra, "Linear attention is (maybe) all you need (to understand transformer optimization)," Oct. 2023, Accessed: May 16, 2024. [Online]. Available: https://arxiv.org/abs/2310.01082v2

[7] "Data-Scientist-Books/Data Preparation for Machine Learning Data Cleaning, Feature Selection, and Data Transforms in Python by Jason Brownlee (z-lib.org).pdf at main · aaaastark/Data-Scientist-Books · GitHub." Accessed: May 16, 2024. [Online]. Available: https://github.com/aaaastark/Data-Scientist-Books/blob/main/Data Preparation for Machine Learning Data Cleaning%2C Feature Selection%2C and Data Transforms in Python by Jason Brownlee (z-lib.org).pdf
[8] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, "Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey," J. Imaging 2022, Vol. 8, Page 205, vol. 8, no. 8, p. 205, Jul. 2022, doi: 10.3390/JIMAGING8080205.
[9] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12901 LNCS, pp. 109–119, 2021, doi: 10.1007/978-3-030-87193-2_11/COVER.
[10] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "SwinBTS: A Method for 3D Multimodal Brain Tumor Segmentation Using Swin Transformer," Brain Sci. 2022, Vol. 12, Page 797, vol. 12, no. 6, p. 797, Jun. 2022, doi: 10.3390/BRAINSCI12060797.
[11] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, "Classification of Brain Tumor from Magnetic Resonance Imaging Using Vision Transformers Ensembling," Curr. Oncol. 2022, Vol. 29, Pages 7498-7511, vol. 29, no. 10, pp. 7498–7511, Oct. 2022, doi: 10.3390/CURRONCOL29100590.
[12] 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 2023, Vol. 13, Page 2094, vol. 13, no. 12, p. 2094, Jun. 2023, doi: 10.3390/DIAGNOSTICS13122094.
[13] E. A. Albadawy, A. Saha, and M. A. Mazurowski, "Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing," Med. Phys., vol. 45, no. 3, pp. 1150–1158, Mar. 2018, doi: 10.1002/MP.12752.
[14] R. A. Zeineldin, M. E. Karar, J. Coburger, C. R. Wirtz, and O. Burgert, "DeepSeg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images," Int. J. Comput. Assist. Radiol. Surg., vol. 15, no. 6, pp. 909–920, Jun. 2020, doi: 10.1007/S11548-020-02186-Z/TABLES/4.
[15] M. I. Sharif, J. P. Li, J. Amin, and A. Sharif, "An improved framework for brain tumor analysis using MRI based on YOLOv2 and convolutional neural network," Complex Intell. Syst., vol. 7, no. 4, pp. 2023–2036, Aug. 2021, doi: 10.1007/S40747-021-00310-3/TABLES/13.
[16] A. A. Asiri et al., "Block-Wise Neural Network for Brain Tumor Identification in Magnetic Resonance Images," Comput. Mater. Contin., vol. 73, no. 3, pp. 5735–5753, Jul. 2022, doi: 10.32604/CMC.2022.031747.
[17] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," ICLR 2021 - 9th Int. Conf. Learn. Represent., Oct. 2020, Accessed: Feb. 21, 2024. [Online]. Available: https://arxiv.org/abs/2010.11929v2