BRAIN TUMOR CLASSIFICATION WITH VISION TRANSFORMERS

Efe Ali ÖZKESİCİ

İzmir Kâtip Çelebi University, Faculty of Natural and Applied Science, Department of Software Engineering, İzmir – Türkiye, (Responsible Author) ORCID: 0000-0002-8716-1063

Assist. Prof. Dr. Osman GÖKALP

Assist. Prof. Dr., İzmir Kâtip Çelebi University, Faculty of Natural and Applied Science, Department of Software Engineering, İzmir – Türkiye, ORCID: 0000-0002-7604-8647

ABSTRACT

Brain tumors stand as complex pathological conditions posing serious threats to human health. The hazardous nature of this disease underscores the importance of early diagnosis and effective treatment. In this context, technological methods offered by modern medicine, particularly advancements in artificial intelligence (AI) and image processing (IP), have made significant strides in the detection of brain tumors.The focal point of this study is to assess the potential of the Vision Transformer (ViT) deep learning model in the context of brain tumor detection. The detection of brain tumors requires precise analysis of complex MRI data, and artificial intelligence has emerged as an effective tool to overcome this challenge. Vision Transformer, distinguished by its ability to process large-scale visual datasets and its learning capacity, is a prominent AI model. This study is designed to evaluate the potential of Vision Transformer in the context of brain tumor detection. Experimental results on a common benchmark dataset show that ViT yields promising results against traditional CNN-based algorithms.

Keywords: Brain Tumor, MRI, Deep Learning, Image Processing, Artificial Intelligence, Vision Transformer

Introduction

Brain tumors are defined as growths of cells in or near the brain. Primary brain tumors originate in the brain, while secondary brain tumors consist of cancer cells spreading from other parts of the body to the brain. Noncancerous or benign brain tumors generally exhibit slow growth, whereas cancerous or malignant tumors tend to grow rapidly. Treatment options depend on the type, size, and location of the tumor; surgery and radiation therapy are commonly employed. Brain tumors can take various forms, with some being cancerous (malignant) and others noncancerous (benign). Symptoms vary depending on the tumor's size and location, encompassing headaches, vision problems, imbalance, and memory issues. Although the cause of brain tumors is often uncertain, factors such as age, race, radiation exposure, and genetic elements are considered risk factors. While there is no preventative measure for brain tumors, individuals at high risk are advised to consider regular screening tests. For several types of brain scans, including magnetic resonance imaging (MRI), well-known machine learning classification techniques have been used (Aldape et al., 2019). Due to the complexity of medical imaging, this topic is still challenging. With the aid of ongoing and upcoming research in this area, we can better understand the course of the disease and create new treatments. Deep learning, a subclass of machine learning, has been used for intelligent systems in many fields, particularly in the processing of medical pictures, and it can handle difficult decision-making tasks (Gassenmaier et al., 2021). Convolutional neural networks have shown promising results for illness diagnosis and organ segmentation (Litjens et al., 2017). CNNs combine the three key phases of a classification job, namely feature extraction, feature selection, and classification, in contrast to conventional machine learning approaches. In automated medical image analysis, convolutional neural networks have achieved major advancements. The Vision Transformer method is based on dividing an image into 16x16 sized pieces and removing their embeddings. Although CNNs have demonstrated success across diverse computer vision tasks and exhibit efficient performance on large-scale datasets, Vision Transformers present distinct advantages in situations where prioritizing global dependencies and fostering contextual understanding are paramount. Vision

Transformers outperform convolutional neural networks (CNN) while using fewer computational resources for training. In this study, we developed a classification model using a vision transformer. The experimental study showed ViT based approaches could be used as a model for automatic detection of brain tumors.

The remainder of this study is organized as follows. In the relevant study section, existing methods and research in the literature on brain tumor detection and classification are reviewed. The next section provides an overview of the entire methodology of this study, including the dataset used and how it was obtained, and how they were used together with ViT. Application details, results and evaluation are discussed in the next section. Results and final thoughts are mentioned in the last section of this article.

Related Works

Kang et al. (2023) presented RCS-YOLO, a rapid and high-accuracy object detector tailored for brain tumor detection. Leveraging the RCS-OSA module, their model exhibited notable performance improvements. The evaluation utilized the 2020 brain tumor dataset (Br35H), demonstrating precision, recall, AP50, AP50:95, FLOPs, and Frames Per Second (FPS) as comparative metrics. The results showcased RCS-YOLO's superiority over state-of-the-art detectors, outperforming YOLOv7 and YOLOv8l in terms of precision, AP50, and FPS, while reducing FLOPs. Additionally, an ablation study emphasized the efficacy of the RCS-OSA module, underscoring its contribution to enhanced accuracy and efficiency.

In another study, Ming Kang (2023) introduced CGFW-YOLOv8 for brain tumor detection, evaluated on the public Br35H dataset. Their model, BGF-YOLO, demonstrated significant improvements over baseline YOLOv8 and competing detectors in terms of precision, recall, and mean average precision (mAP). Ablation studies highlighted the contributions of components like the BRA, GFPN, and the fourth head, showcasing their impact on overall accuracy. Further experiments explored different multiscale feature fusion structures and attention mechanisms, with BGF-YOLO consistently outperforming alternatives. The proposed model's robustness was evident in its choice of regression loss, with CIoU yielding superior results compared to GIoU, DIoU, EIoU, SIoU, and WIoU.

These studies collectively emphasize advancements in object detection techniques tailored specifically for brain tumor detection, with a focus on precision, recall, and overall model efficiency.

Materials and Methods

In this section, the dataset containing brain MRI images is prepared and described for brain tumors. In addition, Vision Transformer explained. After that, how brain tumors can automatically be classified in this study with artificial intelligence is explained with the dataset containing brain MRI images.

Dataset

The dataset used in this study is an open database on Kaggle website (https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset). The dataset contains a total of 3255 images. In this dataset, there are four classes of images. “gomia_tumor”, “meningioma_tumor”, “no_tumor”, “pituitary_tumor”. The main purpose of this dataset is to help researchers develop a model with high accuracy to detect and classify the stage brain tumors.

Data Preparation

In the data preparation phase, libraries such as Tensorflow, Matplotlib, NumPy, and Pandas were installed and used. Afterward, paths to the images were identified and constants and seeds were determined for reproducibility. Then, the dataset was split into %80 train and %20 test batches consisting of 4 classes and determined the number of images for each class was. Images were copied and resampled from the train folder and split randomly with equal proportions in each class for the validation folder.

Data Preprocessing

Data was used in grayscale. Data were rescaled to 20x20 with the help of matplotlib. It is shown in Figure 1.

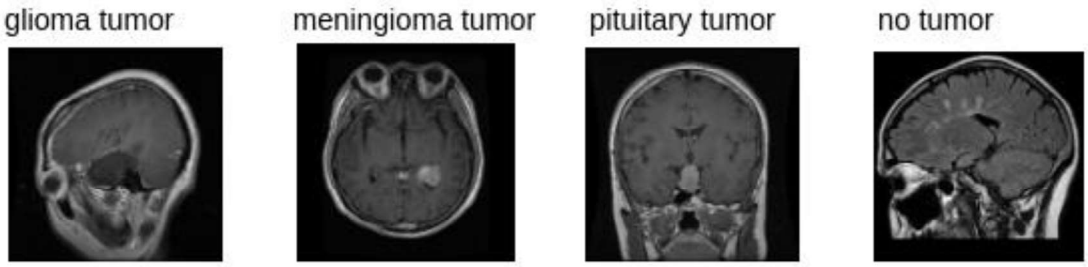

Figure 1. Brain MRI Images

Deep Learning

Deep learning is a subset of machine learning that involves the use of neural networks with multiple layers (deep neural networks) to model and analyze complex patterns in data. It leverages hierarchical representations of features, automatically learning relevant features from raw data without explicit programming. Deep learning algorithms learn to extract hierarchical features through the iterative optimization of model parameters, allowing them to automatically discover intricate patterns and representations in data. The depth of these networks enables them to capture intricate relationships, making deep learning a powerful tool for tasks such as image classification and autonomous decision making. Convolutional Neural Networks (CNNs) are crucial in deep learning for their effectiveness in image-related tasks. By automatically learning hierarchical features from pixel data, CNNs enable the extraction of intricate patterns, revolutionizing image recognition, classification, and other visual data analysis applications.

Vision Transformers

The Vision Transformer (ViT) is a recent breakthrough in computer vision that introduces a paradigm shift from convolutional neural networks (CNNs) to transformer architectures for image classification tasks. While both ViT and traditional CNNs eliminate the need for manually crafted features, ViT distinguishes itself by leveraging a self-attention mechanism to capture global contextual information from the entire image(Dosovitskiy et al., 2020). ViT divides an input image into fixed-size patches, linearly embeds each patch into high-dimensional vectors, and then processes them through a transformer encoder. This enables ViT to efficiently capture long-range dependencies and relationships among image regions. The attention mechanism allows each patch to attend to all other patches, enabling holistic understanding of the image. Pre-trained on large-scale datasets, ViT exhibits remarkable performance on various vision benchmarks, demonstrating its capability to generalize knowledge across diverse visual domains. As a transformer-based architecture, ViT showcases the potential of attention mechanisms beyond natural language processing, marking a significant stride in the fusion of vision and transformer-based models.

Overview of the Developed ViT Model

Patch Encoder using Keras Layer was added to ensure correct encoding of patches used in training. A normalization layer was created from the encoded paths. Multi-head attention layer was created from this created layer. A connection was created between these new layers and patches. Thus, a second layer was created. Normalization was applied to the second layer created and a third layer was created. MLP process was applied to the third layer created. The second and third layers created were added to each other. Normalization and MLP process was applied again to the newly formed layer. Thus, the model is ready for use. You can see the diagram of the model in Figure 2.

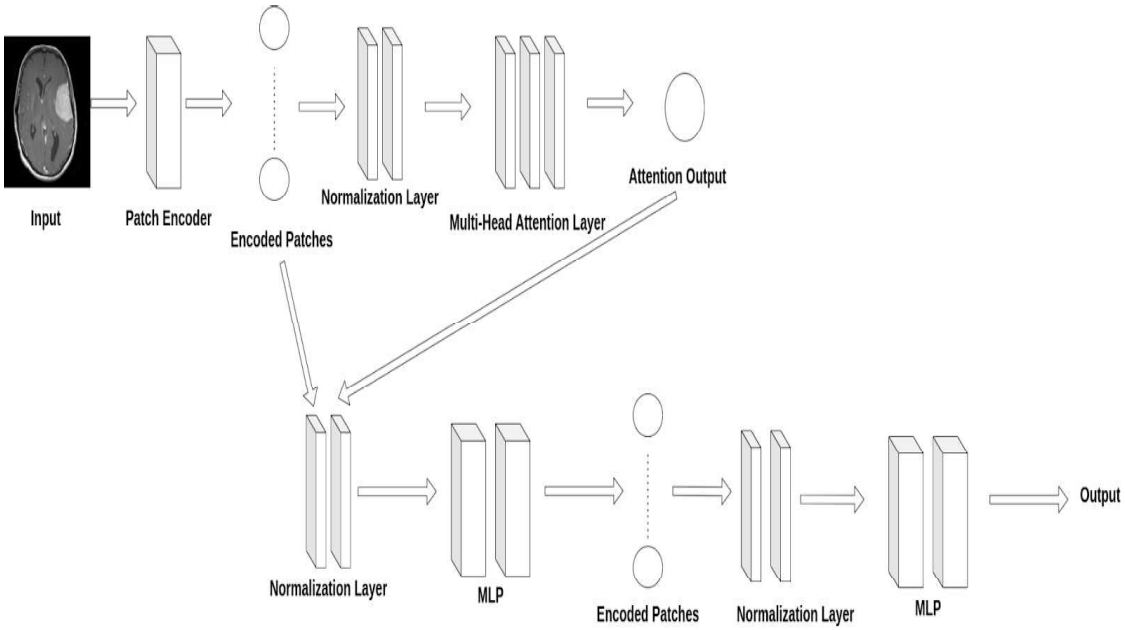

Figure 2. Diagram of ViT Model

Experimental Work

In this section, we evaluate the performance of ViT model with different metrics on dataset containing brain MRI images. During the experiments, the epoch value was set to 10. Data was rescaled to 20x20.

Metrics

The following metrics were used to evaluate the performance of the model described earlier.

1) Accuracy: It refers to the percentage of test samples that were successfully categorized.

2) F1-score: Calculating precision and recall is possible using the confusion matrix. After that, the F1-score is

calculated as the harmonic mean of recall and precision.

3) Precision: Precision, on the other hand, shows how many of the values we estimated as positive are positive.

Performance Evaluation

In this section, the results obtained after experimental work is given and compared against other works that used the same dataset. Table 1 outlines the performance results of three models namely, the "ViT Model," "RCS-YOLO," and "BGF-YOLO." Notably, the "ViT Model" emerges as the standout performer, achieving an accuracy of 0.94, an F1-score of 0.947, and a precision of 0.954. These metrics collectively underscore the exceptional capability of the ViT Model in accurate classification, precision, and maintaining a harmonious balance between precision and recall. The ViT Model's robust performance positions it as a compelling choice for tasks requiring comprehensive understanding and precise predictions. In comparison, while "RCS-YOLO" also demonstrates commendable results in accuracy, F1-score, and precision, and "BGF-YOLO" exhibits promising accuracy and precision figures, the absence of F1-score information for the latter calls for careful consideration. The strong showing of the ViT Model across multiple metrics reinforces its effectiveness, making it a noteworthy candidate for applications demanding nuanced and accurate data analysis.

On the other hand, in the scope of this study, an examination was conducted on the outcomes obtained from a Convolutional Neural Network (CNN) model trained on a comparable dataset. The aforementioned model achieved results of 0.91 accuracy, an F-1 score of 0.85, and precision of 0.90 during the training on the same dataset. However, the results obtained from these experiments appear to be lower compared to the training results achieved with the Vision Transformer (ViT). The superior performance of the ViT model, particularly on intricate medical imaging data, underscores its efficacy. These findings suggest that the ViT model could provide a more sensitive and accurate solution for analogous medical imaging tasks. In this model, changes have been made to the fundamental CNN model. In the 9 layer CNN model, there are 14 stages as well as hidden layers that give us the best result in detecting the tumor. The input image first entered the convolution layer with 32 filters, and then entered the batch normalization and max pooling layers, respectively. After these processes, it entered the convolution layer with 64 filters and was processed again in the batch normalization and max pooling layers. After these operations, the flatten process was applied and finally the dense operations containing 512 and 2 layers were performed and the output was created. The diagram presented in Figure 3 is the projected methodology with a short narration.

The results are shown in Table 1.

Base Model Scores for Brain Tumor Detection	Accuracy	F1-Score	Precision
Trained ViT Model	**0.94**	**0.947**	**0.954**
RCS-YOLO	**0.930**	**0.946**	**0.936**
BGF-YOLO	**0.974**	**-**	**0.919**
Trained CNN Model	**0.91**	**0.85**	**0.90**

Table 1. Performance Evaluation Table With Base Model's and CNN Model

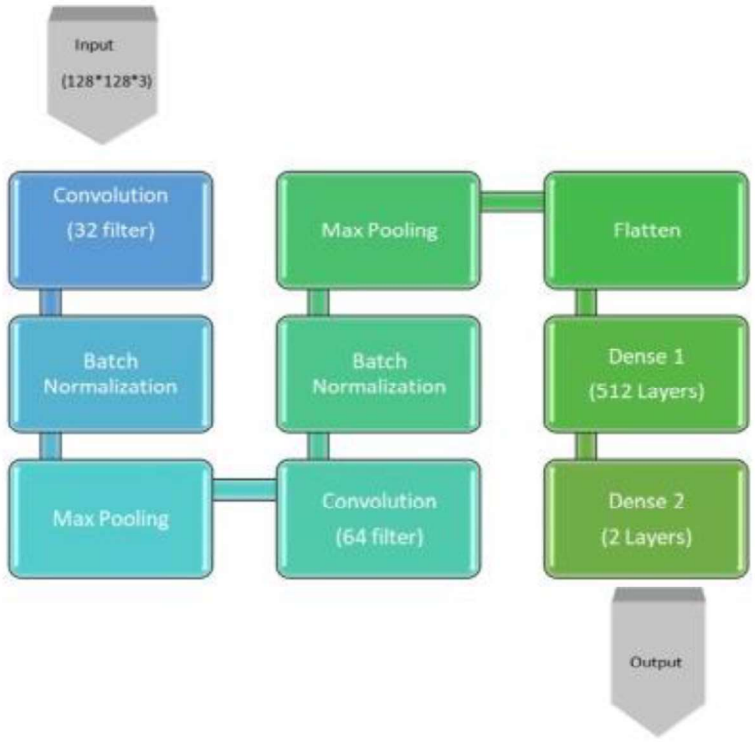

Figure 3. Diagram of Tested CNN Model

Conclusion

In conclusion, this research endeavors to assess the potential of the Vision Transformer (ViT) deep learning model in the domain of brain tumor detection. The ViT model, characterized by its adeptness in processing extensive visual datasets and its capacity for learning, emerges as a noteworthy artificial intelligence paradigm. Notably, ViT excels in highlighting global dependencies and enhancing contextual understanding, outperforming traditional Convolutional Neural Networks (CNNs) by using fewer computational resources.

The experimental findings underscore that ViT-based approaches hold promise for the automatic detection of brain tumors. The ViT Model, showcasing high accuracy, F1 score, and precision, accentuates its proficiency in achieving a harmonious balance between classification, precision, and recall. In comparison with other studies, ViT's robust performance is evident across various metrics.

The success of ViT can be attributed to its unique ability to process an image by dividing it into fixed-size patches and subsequently embedding these patches into high-dimensional vectors. This approach enhances the model's capacity to comprehend long-range dependencies and relationships among different regions of the image. The ViT Model's resilience in dealing with medical imaging complexities positions it as a compelling tool for nuanced and accurate data analysis tasks.

In summary, the ViT Model presents itself as a powerful and effective solution for tasks involving the detection of brain tumors and other complex medical imaging challenges. Future research endeavors may explore additional applications of ViT in diverse medical imaging contexts.

References

Kenneth Aldape, Kevin M. Brindle, Louis Chesler, Rajesh Chopra, Amar Gajjar, Mark R. Gilbert, Nicholas Gottardo, David H. Gutmann, Darren Hargrave, Eric C. Holland, David T. W. Jones, Johanna A. Joyce, Pamela Kearns, Mark W. Kieran, Ingo K. Mellinghoff, Melinda Merchant, Stefan M. Pfister, Steven M. Pollard, Vijay Ramaswamy, Jeremy N. Rich, Giles W. Robinson, David H. Rowitch, John H. Sampson, Michael D. Taylor, Paul Workman and Richard J. Gilbertson 2019. Challenges To Curing Primary Brain Tumors, pp, 3-6

Ingo K. Mellinghoff, M.D., Martin J. van den Bent, M.D., Deborah T. Blumenthal, M.D., Mehdi Touat, M.D., Katherine B. Peters, M.D., Jennifer Clarke, M.D., M.P.H., Joe Mendez, M.D., Shlomit Yust-Katz, M.D., Liam Welsh, M.D., Ph.D., Warren P. Mason, M.D., François Ducray, M.D., Yoshie Umemura 2023. Vorasidenib in IDH1- or IDH2-Mutant Low-Grade Glioma

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A W M van der Laak, Bram van Ginneken, Clara I Sánchez 2017. A Survey On Deep Learning In Medical Image Analysis

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biardc, Aaron Courvillec, Yoshua Bengioc, Chris Palc, Pierre-Marc Jodoina, Hugo Larochellea, 2015. Brain Tumor Segmentation with Deep Neural Networks (BTSWDNN), p, 2-3

Omid Nejati Manzari, Hamid Ahmadabadi, Hossein Kashiani, Shahriar B. Shokouhi, Ahmad Ayatollahi 2023. MedViT: A Robust Vision Transformer For Generalized Medical Image Classification

Tahira Shehzadi, Khurram Azeem Hashmi, Didier Stricker, Muhammad Zeshan Afzal 2023. Object Detection With Transformers: A Review(ODWTR), p, 11-13

Sebastian Gassenmaier, Saif Afat, Dominik Nickel, Mahmoud Mostapha, Judith Herrmann, Ahmed E. Othman 2021. Deep Learning–Accelerated T2-Weighted Imaging of The Prostate: Reduction Of Acquisition Time and Improvement of Image Quality

Arkapravo Chattopadhyay, Mausumi Maitra 2022. MRI-Based Brain Tumour Image Detection Using CNN Based Deep Learning Method