International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307 (Online), Volume-14 Issue-4, September 2024

ViT-BT: Improving MRI Brain Tumor Classification Using Vision Transformer with Transfer Learning

Khawla Hussein Ali

Abstract: ***This paper presents a Vision Transformer designed for classifying brain tumors (ViT-BT), offering a novel methodology to enhance the classification of brain tumor MRI scans through transfer learning with Vision Transformers. Although traditional Convolutional Neural Networks (CNNs) have demonstrated significant capabilities in medical imaging, they often need help to grasp the global contextual information within images. To address this limitation, we utilize Vision Transformers, which excel at capturing long-range dependencies due to their self-attention mechanism. In the case of ViT-BT, the Vision Transformer model undergoes pre-training followed by fine-tuning on specific MRI brain tumor datasets, thereby improving its capability to classify various brain tumor types. Experimental results indicate that ViT-BT outperforms other CNN-based methods, delivering superior accuracy and resilience. Evaluations were performed using the BraTS 2023 dataset, comprising multi-modal MRI images of brain tumors, including T1-weighted, T2-weighted, T1CE, and Flair sequences. The ViT-BT model showcased remarkable performance, achieving precision, recall, F1-score, and accuracy rates of 97%, 99%, 99.41%, and 98.17%, respectively. This advancement is anticipated to significantly enhance diagnostic accuracy in clinical settings, ultimately leading to improved patient outcomes. The research underscores the potential of transfer learning with Vision Transformers in medical imaging as a promising avenue for future exploration across various medical domains.***

Keywords: Deep learning, Vision Transformer (ViT), VGG16, EfficientNet-B7, Transfer Learning.

I. INTRODUCTION

Notably, the condition of brain tumors is among the most frequent and severe diseases that people suffer from in the present day. MRI is a common technique used to diagnose brain tumors [1-3]. Even though brain tumors may manifest in people of any age, it is highly identified the nature of an illness in offspring and older adults [4]. General malaise, sensitive development, and sickness are essentially the most frequent troubles that are brain tumors and headaches. Of all the symptoms, headaches are the most persistent, and they are mostly reported to be painless or sharp. Sickness is defined as disease, forgetting, and fascicular gyrations.

Manuscript received on 19 August 2024 | Revised Manuscript received on 28 August 2024 | Manuscript Accepted on 15 September 2024 | Manuscript published on 30 September 2024.
***Correspondence Author(s)
Khawla Hussein Ali*, Department of Computer Science, University of Basrah, Iraq. E-mail : khawla.ali@uobasrah.edu.iq, ORCID ID: 0000-0002-5569-4578

Other psychological changes that are found in patients with brain tumors include memory, confederation, and analysis difficulty. Two other symptoms associated with brain tumors include vision and hearing impairment, limb agnosia, and speech disorders [5]. Experts utilize several aspects to categorize and, in turn, diagnose brain tumors. Such factors as location, size, and certain characteristics that may be seen on imaging play this role [6]. Meningioma is one of the varieties of brain tumors that originate from the brain's meninges, which are described as bleary tissues. Gliomas tumors develop from glial cells, and glioblastomas develop from the brain. Gliomas and glioblastomas are the same because both are cancerous tumors that grow in the brain [7]. Another tumor category is the pituitary, located in one of the essential glands in the head. To the pituitary gland, other glands in the body can be explored conveniently. Research has also shown that specialists can effectively identify and manage any disease associated with brain tumors [8] concerning the mentioned characteristics of these diseases. Chemotherapy, surgery, and radiation therapy are considered the ordinary means of treating brain tumors: ultra-violet light, chemical agents, and operative. The occurrence of brain tumors affects both the patients and their families. Consequently, it is only logical that early diagnosis of the diseases will lead to a better prognosis [9]. The following are some of the imaging methods useful in the diagnosis of brain tumors: Of all these models, MRI is the most popular, and it is an abbreviation for magnetic resonance imaging. MRI helps identify and detect Brain tumors. Employing magnetic fields and radio waves [10]. Another modal is computed tomography, also referred to as the CT scan. This particular model uses images made with the help of X-rays to detect and identify the presence of any brain tumors. Brain tumors could also be diagnosed with PET, commonly known as Positron Emission Tomography. In this case, imaging is done by injecting the body with a radioactive substance that circulates in the blood. Among the proposed models of surgical procedures, the request mechanism is the most suitable when it comes to the detection of brain tumors. Histological confirmation is made by a biopsy whereby a small portion of the tumor is subjected to a microscope to help define the description of the brain tumor. All these imaging models are helpful for the detection of brain tumors [11]. However, as to the above-mentioned classical forecasting methods, they also have their shortcomings. Such imaging models are expensive and time-consuming. The factor that makes them functional can pose a challenge to patients who often need to leave the wards for check-out scans [12].

Also, the precision of these imaging models can be impressive due to the tumor's amount and area and the presence of surrounding materials. This classification model has other assets of false results symbolized by the confusion matrix. This is because correctly identifying the disorder is a capital exercise, leading to either misdiagnosis or delayed diagnosis [13]. Furthermore, it has been found that large tumor sizes and locations influence the reliability of these imaging models, as well as the existence of the surrounding tissues. There are probabilities of false results, as depicted in the confusion matrix of the developed model. This can result in wrong conclusions and causative treatment being given at a bad time. Some machines and deep learning models developed to find and diagnose brain tumors have also been exploited to eliminate the issues above. Data is fed to the model from the patient's MRI scans/ images, and the model is trained using algorithms. The model then uses this input image, or another, to predict the presence or appearance of brain tumors [14]. In classification and segmentation, SVM can identify and detect Brain tumors efficiently. Random forest (RF) is the ensemble learning algorithm that can include categorical analysis, known as classification, and the analysis of prediction, known as regression, for detecting tumors. CNN again refers to the convolutional neural network utilized in detecting brain tumors and is categorized under deep learning. CNN is under the category of image classification. This algorithm can extract the features from the brain images without necessary intervention from a human expert, and it sorts the images depending on these specific features and the ability to detect the presence of brain tumors [15]. In [16], the paper used a pre-menu transfer learning, the VGG-16 model, as a medium to detect several kinds of brain tumors. For this paper, the dataset that has been considered is the CE-MRI dataset, which contains MRI images of brain tumors belonging to four special categories. The dataset used in the research is 233 patients' MRI scans and images, a total of 3,064, which were utilized initially for training the model and validating the output predetermined from the model. Hypothesis H 1 was supported by the study findings, thus showing that the proposed methodology was sufficient, as seen by the 94% accuracy in sorting the brain tumor images into four categories. In [17], the authors learned a CNN method with one special group's data. They validated it on data collected from two other groups to check whether different data capturing and analysis modes from the research institution would impact the algorithm's performance. The authors employed the dataset from the TCIA data

The results revealed that CNN's performance was comparatively better at the initial stage when trained on image data or scans of similar groupings or institutions than when trained on data from one institution and tested on another clinic. In particular, an empirical analysis of the research hypothesis was carried out based on the null hypothesis that the DSC score for the clinic equals 0. 76 ± 0. 12. The paper [18] introduces a deep learning framework to diagnose tumors using MRI information. The authors presented a deep 3D CNN architecture named BraTS-Net, which includes the organization of 2D and 3D convolutions and a contention of modules to enhance p achievement. The dataset employed in this work was the BraTS 2013 dataset, which contained MRI images and scan data and recorded patients with necrosis, edema, and LGG. In the proposed work, compared with the SVM models, RFs defined that the BraTS Net gets a significantly better score than the models, and the gross DSC score is 0. 88. However, the paper does not comment on the applicability of CNN to other datasets to the best of the author's knowledge. The authors of [19, 20] employed a CNN architecture containing three filtered layers and two dense layers for the segmentation of MRI of tumor-diagnosed patients with the help of the BAT algorithm. The authors' datasets were MRI image data of glioma patients obtained from BraTS 2015. The model's specificity and sensitivity were acceptable, rated at 87% and 90%, respectively, and accuracy was 92%. Nevertheless, the authors fail to compare the described model with professional advisors, which can shed light on some of the model's potential shortcomings. The authors developed a deep neural network framework called DeepSeg [21], designed to segment brain tumors into four categories: edema, non-enhancing tumor, enhancing tumor, and necrosis.

After evaluating the performance results, it was observed that the performance of CNN was relatively better at the initial steps when trained on either image data or scans of similar grouping or institution than when trained on one institution's data and tested on data from another clinic. In particular, an empirical examination of the research hypothesis was done concerning the null hypothesis posited that the clinic's DSC is equal to zero. 76 ± 0. 12. The paper proposes A deep learning architecture for tumor diagnosis from MRI information [18]. The authors proposed a novel abysmal 3D CNN architecture named BraTS-Net, which contains the structuring of 2D and 3D convolutions and the existence of the set of modules to raise p achievement. The data set used in this work was BraTS 2013, comprised of MRI images and scan data that annotated the patients with necrosis, edema, and LGG. Finally, in the proposed work, RFs compared that BraTS Net achieved much better performance than the models, where the gross DSC score of the models is 0. 88. To the best of the authors' knowledge, the paper does not express an opinion on the generalization of the CNN to other datasets. [19] used a CNN architecture consisting of 3-filtered layers and two dense layers for the segmentation of MRI of tumor-diagnosed patients and used the BAT algorithm for the same. The dataset used by the authors was MRI image data of glioma patients, which the authors retrieved from BraTS 2015. The specificity and sensitivity of this model can be considered sufficient and amounted to 87 % and 90% respectively, accuracy of this model was 92 %. However, the authors do not proceed with the implementation of the comparison of the described model with professional advisors to emphasize some of the possible drawbacks of the model. The authors developed a deep neural network framework called DeepSeg [21, 22], designed to segment brain tumors into four categories. Secondly, the categorization used is edema, non-enhancing tumors, enhancing tumors, and necrosis. The authors in another paper [27] described the challenges in multimodal brain tumor detection and segmentation issues in MRI, including the problems of variability and versatility of the tumor tissue through a Transformer network capable of capturing the complexities between the modality.

Retrieval Number: 100.1/ijsce.D364414040924
DOI: 10.35940/ijsce.D3644.14040924
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The introduced method is called Transbts and includes two steps. This is done in the preprocessing part, where input MRI images are computed to get one multi-scale feature map. The Transformer network is also adopted to split the feature maps and send them to the segmentation part. The authors apply several ML models to analyze the model they proposed. The results revealed that when using the whole tumor, the mean Dice coefficient achieved was 90% (BraTS 2019) 90. Three weeks for the ISNT MT and 09% for the cancer (BRATS 2020). Another strength of this paper is that the authors used two rich yet easily accessible datasets, enhancing the results' dependability and generalization. However, the authors shall explain the time taken in computational resources in training and reasoning, which could be a significant limitation in applying and implementing the model in its use as a reliable medicine model. These machine and deep learning mechanisms are beneficial for identifying brain tumors. However, they have these drawbacks. These models' output comes up on the clean, quality data and amount and size of the data. It is equally important to note that these models may not detect other more complicated types of tumors that are a special and relatively rare classification of brain tumors. Different parameters can include the feature selection, and other parameters can also influence the accuracy of these algorithms [24, 25].

With this shortcoming in mind, this weakness publicized the need for a better, more accurate technique for diagnosing brain tumors. Vision Transformer can affect present restrictions, a deep-learning model for promising prospects. Vision Transformer using a neural network framework is a distinct strategy for computer vision tasks [32-34]. Through self-attention techniques, the ViT model pushes and detects complex tumors by forming and recognizing splits from the input image. This model learns the normal healthy tissue and the abnormal malignant cancerous tissue during training by reducing a specifically identified loss of function. It is obtained at the testing phase as a probability map indicating information about areas of the brain with tumors [25, 26]. This paper addresses the problem statement: Can a vision transformer model label and locate brain tumors in medical images efficiently? The assumption is that vision transformer models trained and learned on MRI scans can conclude high efficiency in identifying and detecting brain tumors and that the current models for brain tumors can be exceeded. In contrast, in the classical models, based on the convolutional neural network for quantitative assessment of 3D relations between the pixels in the brain tumor images, the ViT extracts the global relation between the pixels using a self-attention mechanism. Vision Transformer learns the features of the MRI scans of brain tumors and can determine the tumors. Some of the developed datasets of brain tumors are huge medical images and other clinical data that can be applied to machines and deep learning algorithms. By using these datasets, clinicians and researchers can identify, outline, and analyze the presence of brain tumors. A few of these brain tumor datasets are the datasets of MRI images of the most frequent yet invasive kind of tumor referred to as BraTS (The Brain Tumor Segmentation Challenge Dataset). The collected data contains MRI scans and associated scientific data of over two hundred patient cases. BraTS is widely utilized concerning the identification, detection, classification, and segmentation problems of brain tumors [27]

Low-grade glioma (LGG) has better progress than other brain tumors because it grows slowly. Other datasets are available with glioblastoma MRI scans with certain transformations (T1 image, T2 image, T1CE image, and Flair image) [28]. Another dataset benchmark is TCGA-GBM, which includes MRI and genomics data on glioblastoma. In addition to the genomic data, MRI scans, and other patient clinical details, the CPTACGBM also contains proteomics of tumor patients [29,30]. All the necessitated datasets described above are used to train various machine and deep learning models. They then use these models to evaluate input images to detect, classify, and segment brain tumors. The paper contributions are:

- Estimate a correct model of performance for diagnosing brain tumors from image.
- In this case, the Vision Transformer (ViT) model is a technique that will be applied to conclude that it offers higher accuracy and shorter times in terms of reasoning for tumor identification.
- Discuss how to encourage the outcome's performance using transfer learning such as VGG16, Efficient-NetB7, and data augmentation.
- Supply a dependable and autonomous apparatus for primary identification and analysis of the brain tumors, thus decreasing the high false positives and enhancing the medical system

Nonetheless, the classification of the actual brain tumor MRI images that are accurate, authentic, and active is a big problem due to the significant presence and area of variability of tumors. The early techniques for classifying brain tumor MRI images were the conventional techniques that involved hand-engineered features and machine learning algorithms that were defined by their capacity to learn high-order features and variations in the data set. Recent progress in deep learning models has provided promising performance in the medical image classification problem, including brain tumor classification. Specifically, the structure of the transformer has gained significant attention in computer vision tasks because of its ability to capture global contexts and collect spatial relations between the image's features. Nonetheless, the use of transformers in medical image classification is still a somewhat active research area today [31]. The general transfer learning approach has been implemented in the various classification of MRI images. In [4], the paper gives general information about CNN concerning brain image analyses and mentions different architecture and techniques for feature extraction. Still, it does not account for specific issues related to deep learning for MRI image classification.

In [48], the paper examines the use of transfer learning through pre-trained CNNs such as VGG16 & ResNet for feature extraction and fine-tuning for MRI datasets. The paper can fail to explain some shortcomings, including domain shift or data bias.

In the context of transfer learning for ultrasound image classification, the paper's focus is somewhat narrower, mainly because it is already established that results derived from one imaging modality may not be directly portable to another modality. In contrast, the paper does not dwell on the complex forms of transfer learning or other complex architectures that could be useful for improving the classification of MRI images; there is certainly much more research work on that subject beyond the scope.

In [3], the paper discusses deep residual connections through which Res-net enables the training of intense networks. As opposed to work considering the peculiarities of MI images, including image resolution, noise, or anatomical variations in the part dedicated to the proposed solution, the paper is oriented towards ResNet architecture and its performance on benchmark datasets, except for the parts concentrated on segmentation and survival prediction.

The proposed work in this paper is a vision transformer-based approach given the acronym ViT-BT for breast tumor MRI image classification. The specific technique is a fine-tuning approach applied to a pre-trained vision transformer model, with outstanding results in computer vision tasks. The approach builds on vision transformers' capability to capture relatives between elements in images and understand the images globally. The rest of the paper is organized as follows: Section 2 describes the materials and methodology used for the proposed method, Section 3 describes the detailed results, and then the paper is concluded in Section 4.

II. MATERIALS AND METHODS

A. Dataset

In this work, we utilized the BraTS-2023 dataset [23], which is a retrospective database of glioma MP-MRI acquired at different sites under clinical protocols using different scanners and different imaging sequences, so this cohort represents massive variability in terms of image quality as it imitates the variability of the current clinical practice. Every tumor sub-region was confirmed by two independent neuroradiologists who concluded with expert ground truth annotations. Each MRI scan in the dataset contains four different modalities: T1 MRI, T2 MRI, T1 MRI with contrast enhancement, and FLAIR MRI. The dataset is divided into three parts: the ratio of training, validating, and testing as 80:10:10 correspondingly. The described MRI scan dataset contains ground-truth labels for segmentation and classification tasks regarding each MRI scan. The ground-truth labels against which to compare the TTLES algorithm are the tumor core, the enhancing tumor, and the whole tumor. The necrotic and non-enhancing tumor core area consists of the necrotic region in the tumor mass, the enhancing tumor area consists of the part of the tumor that enhances after the injection of contrast medium, and the whole tumor area is the union of the tumor core area and the enhancing tumor area as depicted in fig 1.

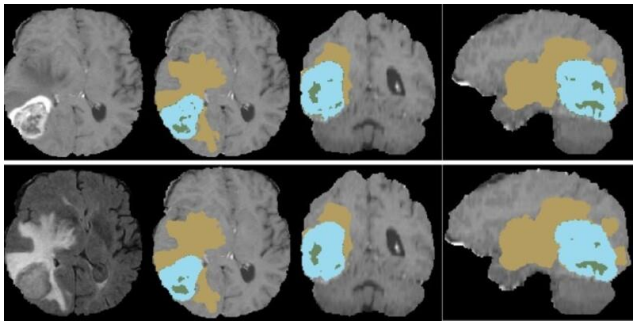

Fig. 1: Sample of Bra TS Dataset

The BraTS 2023 dataset also includes information such as patient age, sex, tumor location, and histology. This information can be used to explore the relationship between these factors and the characteristics of brain tumors. The BraTS dataset is widely used for developing and evaluating brain tumor segmentation and classification algorithms. The dataset has been used in numerous studies and competitions and has contributed to significant advances in medical image analysis.

B. Deep Learning- Transfer Learning

i. VGG16

VGG is an architecture of deep convolutional neural network formed by the Visual Geometry Group (VGG) of the University of Oxford in 2014. It was designed to compete in the ILSVRC and was very successful; the network won the localization and classification of the ImageNet dataset [48]. The VGG16 network comprises 16 layers of conducting convolution and complete connection. The 1st and 13th layers are the convolutional layers, and the 14th and 16th layers are the fully connected layers. The convolutional layers use small 3x3 filters and are arranged so that multiple of them are connected, and therefore, it has a deep structure. By the time the information has passed through the first couple of convolutional layers, several filters in the convolutional layers process input images and detect edges, corners, or even textures. There are fully connected layers at the end of the network responsible for classifying the output of the convolutional layers to the class in the ImageNet database. VGG-16 has more than 138 million parameters; thus, it is a robust architecture to work with images for classification, as in Fig. 2.

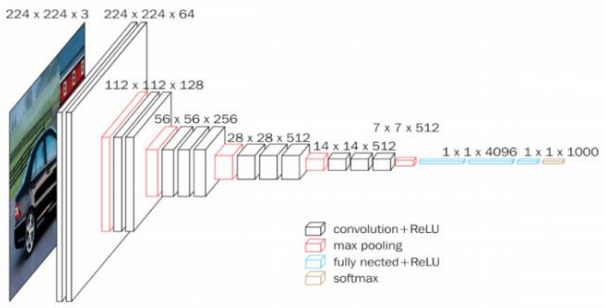

Fig. 2: VGG Architecture

Retrieval Number: 100.1/ijsce.D364414040924
DOI: 10.35940/ijsce.D3644.14040924
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

ii. EfficientNet-B7

The model scaling also depends on the baseline network; that influence was described similarly. Therefore, to fine-tune the results even more, The preceding creates the usage of the MBConv, the mobile inverted bottleneck convolution [8]. Technically, there is no such model as EfficientNet-B7. The EfficientNet-B7 architecture comprises multiple models, EfficientNet-B0 to EfficientNet-B7, which differ in the depth of the model. EfficientNet-B7 can be defined as an array of deep neural networks that Google View proposed to research in 2019. The architecture is meant to deliver high competence on image classification tasks using considerably fewer parameters and computational power than other well-known architectures like ResNet and Inception [40]. From the EfficientNetB-7 models, the compound scaling approach scales depth, width, and the network's resolution uniformly. This enables the model to achieve the desired aims of solving the trade-off where more parameters could mean better accuracy. Still, it could take a long time to compute, resulting in fewer parameters. As stated earlier, B0 is the leanest and most efficient family variant, but B7 is the biggest and most powerful. EfficientNet-B7, as in Fig. 3, which has over 66 million parameters, shows the state-of-art on various image classification datasets. Nevertheless, the training and deployment of such a big model entails considerable computational power and skills. In conclusion, Efficient-Net is one of the best architectures that you can use for image classification problems, and its several models are implemented in many practical applications.

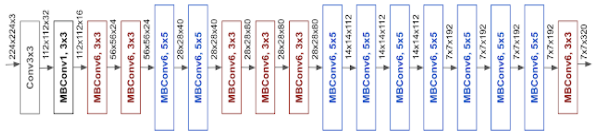

Fig. 3: Architecture of EfficientNet-B7

C. Methodology of Vision Transformers

In this section, we discuss the experimental setup for our study, the metrics we have used to evaluate the implemented models, and the implemented models themselves.

i. Experimental Setup

The literature for our study is centered on comparing two deep-learning models for image classification. In this study, we have implemented and trained the models using the same approach.

ii. Data Preparation

Obtain the tumor brain MRI dataset and then perform the data augmentation, normalization, and resizing processes. Divide some of the obtained data into the training data set, the validation data set, and the testing data set.
1. In the feature extraction stage, a given input image is divided into small, non-overlapping, fixed-sized regions. Each region takes a section of the image as a local area where processing occurs. Then, each patch is linearly embedded into a vector space of a lower dimension than the original data. This step maps the patch's pixel values into a smaller vector space.
2. Positional Encoding: Include the positional encoding for the patch embedding sequences to carry information about the positions of patches in image parts.
3. Transformer Encoder: Several transformer encoder layers are used for the patch sequences to gather the context between the image parts.
4. Classification Head: Feed the transformer encoder to include a classification head to give an output for the yes/no presence of a brain tumor in the MRI data.
5. Training and Evaluation: Train the Vision Transformer model on the training set and then assess the experiment's success based on the chosen evaluation measures, including accuracy, precision, recall, F1 score, AUC-ROC, and AUC-PR in the validation and testing sets.
6. Hyper Parameter Tuning: Hyperparameters of the Vision Transformer model include the learning rate and the number of layers for the Vision Transformer, image patch size, dropout rate, the number of attention heads, and more. Table 2 evaluates the Vision Transformer model results against other models as a case study of tumor brain MRI data. It compares deep learning models such as CNNs and hybrid models that combine CNNs with RNNs or attention mechanisms.

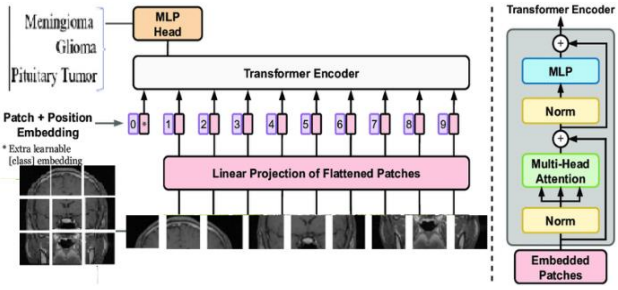

Fig. 4: ViT-BT Architecture for MRI Images

The algorithm of the ViT-BT is as follows:

	Algorithm of Proposed ViT-BT
1-	Combine VGG-16 and EfficientNet-B7 with Vision Transformer: Integrate the features extracted from the transfer learning with the Vision Transformer. This can involve concatenating the features or using them as additional inputs to the ViT model.
2-	Fine-tune the Combined Model**:** Fine-tune the combined model on the MRI image classification task.
3-	Divide the brain tumor MRI image into patches and flatten it as a vector
4-	A sequence of patches is mapped with a trainable linear projection
5-	A learnable class embedding Z_{class} is before the sequence of embedded image patches
6-	The patch embedding is finally followed with 1-D positional embedding E_{pos}.
7-	The sequence of embedding vectors is: $Z_o = [X_{class}; x_p^1; x_p^2; \dots\dots; x_p^N] + E_{pos}$
8-	Compute classification by feeding Z_o at the encoder
9-	Take Z_{class} at the layer Lth of the encoder output and feed it to a classification head.

As of algorithm and architecture in Fig. 4, in the first step, an input image of shape (height, width, channels) is embedded into a feature in a single vector of shape (n+1, d), following an arrangement of transformations. This corresponds to Equation (1):

$$z_o = [X_{class};\ x_p^1;\ x_p^2; \ldots\ldots;\ x_p^N + E_{pos} \quad \ldots.(1)$$

1. The image is partitioned or segmented into n equal small squares of size predetermined as p, and each segment is of the shape (p, p, c).
2. The patches are flattened, which means that the number of vectors representing the n line is n, which has the shape (1, p^2*c).
3. The flattened patches are then multiplied by a trainable embedding tensor of size p^2*cxd. Here, x is a hyperparameter of the architecture and is adapted for different components, and d is the fixed dimension of most parts of the architecture. The upshot is N embedded patches of shape (1,d).
4. A learnable token of shape (1, d) is concatenated to the sequence of patch embeddings, where d is the dimension of patch embedding. This token is from the BERT paper, and only the last representation, the output of the transformer L, is passed through the classification layers. This is a total of the representations of the patches, which is quite intuitive.
5. A trainable positional embedding tensor, E_{pos}, with the same shape (n+1, d), is added to the concatenation of the above projection sequence. Getting back into the flow of this tensor, this tensor learns 1D positional information for each patch to add a spatial representation of each patch in the sequence. The final output fed to the stacked transformer encoders is represented by z_o. The L-stacked encoders constitute the second part of the mentioned architecture. Every transformer receives features in the form of an (n+1, d) tensor and outputs a similar tensor. In the second step, the network learns higher-level features from the embedded patches with the help of the stack of L transformer encoders.

MHA and a 2-layer MLP are in the encoder component; it includes layer normalization and residual connections for added benefit. Layer normalization helps make the hidden state more stable and accelerates the number of training steps. It is computed by scaling with the mean and std of the training example (as opposed to the batch norm, where this is done across features). The resultant features are then scaled by a factor and shifted by another factor learned during the training phase. Residual connections provide the gradients with an alternative route, thus addressing gradients vanishing in networks with deep architectures. In this component, the trainable weights would be restricted only to the inside of the MHA mechanism and the MLP weights. Since the MLP has two layers (hidden and output), there will be two weight matrices: The Weight matrix W_h has a shape of (d, d_{mlp}) while the output Weight matrix W_o has a shape of (d_{mlp}, d). It should be noted that the MHA step included in each of the L-stacked transformers is analogous to Equations (2). (3), (4) and (5).

$$[q, k, v] = zUqkv \quad \ldots.. (2)$$

$$A = \text{softmax}\ (qk^T / \sqrt{D_h}) \quad \ldots. (3)$$

$$SA(z) = Av \quad \ldots.. (4)$$

$$MSA(z) = [SA_1(z); SA_2(z); \ldots\ldots\ldots SA_k(z)\ ]\ U_{msa} \quad \ldots..(5)$$

Subsequently, Qi, Ki, and Vi denote the input projection in 3 sub-spaces of the FA3 module. Every line in Q is a learned projection of the patch, and lines in K are other patches to compare with Q. V and K are learned to quantify the importance, or weights, of features in V to compute the final "attention." The self-attention is the product between A and v, which has the shape of (n+1, d_h). The element on row i and column j is the weighted average of the feature j by the pdf on line i in A. The self-attention matrices are stacked on the second dimension to build an (n + 1, d) tensor, passed to a single-layer preprocessing, multiplying it by a trainable (d x d) tensor. This linear layer is crucial because it enables features to be learned from all the heads as an aggregate.

III. RESULTS AND COMPARISON

The main idea of this paper was to propose the method of Brain Tumors, Vision Transformers classification and detection (BT-ViT). The motivation was viewing the urgent necessity for the precise and non-hazardous identification of brain tumor types to facilitate the decision in medical diagnosis and medicine planning. Expressing the potential of Vision Transformers, it was work directed at obtaining high accuracy in forecasting and classifying various forms of brain tumors: T1, T2, T1CE, and Flair. The work aimed to test the effectiveness of the identified ViT-BT model by comparing it with the existing models while searching for the most effective solution in the field of brain tumor detection in the given field of medical imaging. The dataset used in this study comprised 2040 MRI scans, each belonging to one of four classes: T1, T2, T1CE, and Flair. The given dataset was divided into 80:20 fashions for training and validation. The quantity of the first collection was 1632 images, and the amount of the second one was about 408 images. The ViT-BT model was used, and the input images and preprocessing steps were performed using various methods such as cleaning and enhancement images. The ViT model was evaluated with the help of accuracy, precision, recall, and F1 score to assess the efficiency of the proposed approach for classifying brain tumor images. The model attained great accuracy in validation, which was a significant representation of 98 percent. Among respondents, 17% affirmed their ability to correctly distinguish brain tumors, and 27% of the company's patch size was valued, as shown in Table 1.

Table 1. Values of Patch Size

Model Architecture	Patch size	No. layers	Hidden size	MLP size	Learning Rate	Parameters
ViT	18	14	788	5072	0.0001	88M

The initialization weights come from transfer learning using supervised Image Net pre-trained weights. Learning Rate (LR) starts with 1e – 4 and is then divided by two each time the model iterates 100, resulting in better accuracy. The optimizers used were Adaptive moment estimates (ADAM) [9]. The type of activation function used is the Rectified Linear Unit (ReLU) [7]. Evaluation of a classification model in classification tasks is based on specific evaluation metrics to measure the ability or efficiency of a given machine learning model to predict class labels of input instances. Here are some commonly used evaluation metrics for classification tasks [35- 38]: Here are some widely used evaluation metrics for classification tasks:

1. **Accuracy:** The percentage of correctly classified cases out of the total cases. Formula: The Accuracy Formula is given by TP + TN / TP + TN + FP + FN where TP is the True Positives, and TN is the Total Instances.

2. **Precision:** The extent to which the number of instances is definitively positive out of the total number of cases predicted as positive.

Formula: TP, where TP is several true positives, ought to be equal to TP / (TP + FP).

3. Recall: The percentage of specific positives to the overall sum of particular positives and negatives.

Formula: True Positive /(True Positive + False Negative)

4. F1 Score: Precision divided by the recall." Formula: 2 * (P * R) / (P + R) The AUC-ROC score is between 0 and 1 and indicates that it has the highest accuracy of all the negative/positive cases identified and studied by the classifier. 5 denotes chances being taken, while 1 denotes perfect classification.

5. Area Under the Precision-Recall Curve (AUC-PR): An evaluation metric for binary classifiers that computes the area of the precision-recall curve. The precision-recall curve is a graph that plots precision on the Y-axis concerning recall that we get while using different thresholds. Therefore, the correct evaluation metrics must be selected depending on the task and the specific circumstances of the problem. Some of the evaluation measures might be more appropriate for assessing the models considered in an imbalanced data context, while others can be more appropriate for Multi-classification problems. Here is an overview of the steps to implement a Vision Transformer model on tumor brain MRI data: For the smooth procedure of the organization's implementation, it is necessary to select the right metric, preprocess the data, and fine-tune the model's hyperparameters. Further, we have to compare the results of Vision Transformer with other state-of-the-art deep learning models and analyze them to detect the most efficient model for this particular task. In conclusion, using a Vision Transformer in tumor brain MRI data provides an excellent opportunity for the development of medical imaging and the improvement of patients' outcomes. Despite several resumes and occurrences, the optimum achievable fine in the ViT models was noted evidently in Table 2, which presents the result of VGG16, EfficientNet-B7, and ViT-

Table 2. Evaluation Metrics

Model Architecture	Recall	Precision	F1-score
VGG16	0.95	0.94	0.95
Efficient B7	0.97	0.95	0.96
Proposed ViT-BT	0.98	0.97	0.98

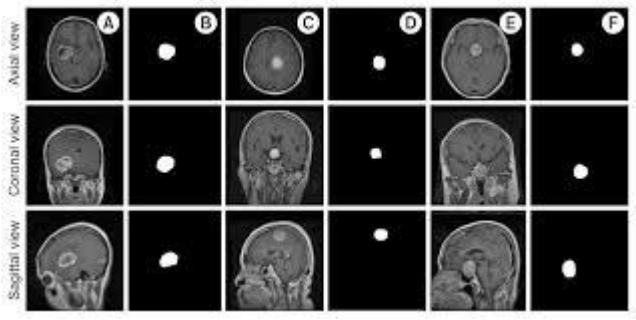

Fig. 5: Output Images of Proposed Model

In Fig. 5, representative results of the vision transformer for brain tumor classification of 'ViT-BT' on the image from the BraTS 2023 dataset. Moreover, it can be concluded that the model works properly with training and testing data and can classify each region in the image correctly. In this paper, it is evident that the self-supervised ViT-BT could obtain the total best achievement. As per the effort and time spent on technician annotation by experts, one cannot get hold of a large set of labeled medical images; however, it may be easy to acquire many unlabeled photos.

Fig. 6: Confusion Matrix

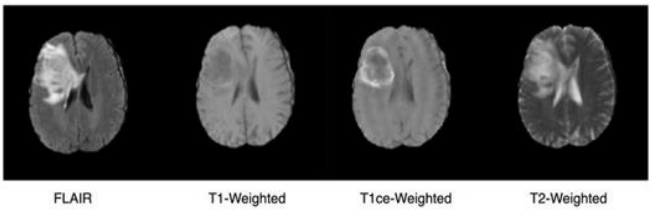

Fig. 7: Correct Prediction Result of Proposed ViT-BT

The confusion matrix discussed in Fig. 6 was employed to gain more insight into the model's performance by showing the correctly classified and misclassified instances. For the 408 images in the validation set, the model accurately classified 393 and misclassified 15. In addition, the prediction accuracy of the proposed model is illustrated in Fig. 7, which shows that the four classes of brain tumors are 98%. An accuracy equal to 17% was obtained for each class. Cohort analysis of the results presents high predictive accuracy; hence, the above confusion matrix highlights the model's suitability in diagnosing all four tumor types.

In the case of <aspects to target from the list>, it demonstrates how using the vision transformer model offers reliability in identifying the correct labels for images of MRI scans.

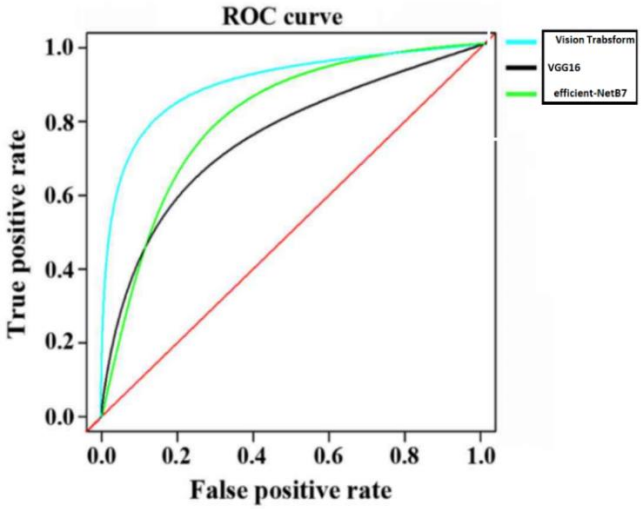

Fig. 8: AUC (Area Under Curve)

In the case of the classification models, the Area Under Curve (AUC) is used, illustrated in Fig. 8. It is utilized in instances where the sensitivity and specificity of the model, which are expressed through actual positive rate and false positive rate correspondingly, matter. When the AUC has high accuracy, it shows that the model performs well in categorizing the tumor instances into different classes as it is highly efficient in ranking them [44- 47]. This infers that a higher value of AUC will mean that the model has a high actual positive rate and, at the same time, it has a low positive rate. In other words, the model effectively captures the relevant group of positives and the tumor instances and has a low false positive rate. It also proves to be specific in the ability to distinguish between various classes of the tumor. It has been shown that machine learning algorithms can make a rather precise prognosis of an individual's prognosis and lethal outcome while retaining a high clinical relevance overall accuracy and true-positive rate [39-43].

Table 3. Evaluating Model in Tumor Classification

Table 3: Evaluating Model

	Precision	Recall	F1-Score	Support
T_1-Weighted	0.99	0.99	0.99	109
T_2-Weighted	0.95	1.00	0.97	82
T_1CE	0.99	0.95	0.97	104
Flair	1.00	0.99	0.99	113
Accuracy			0.98	408

The results of the classification analysis presented more specific information about the model's effectiveness in each tumor class, which is given in Table 3. Finally, the precision for the Flair tumor class was 1. 00, with a recall of 0. 99 and an F1 score of 0. 99, supported by 113 images. Regarding the T2 class, the model has the following classification performance: precision of 0. 95, recall of 1. 00, recall of .88, and F1 score of 0. 33 with the help of 45 pictures, and group C scored 97 with 82 pictures. The model's performance in classifying the T_1CE tumor class yielded a precision of 0. 99, recall of 0. 95, precision of 0.95, and F1 score of 0. 97% of the participants supported the findings with the help of 104 images. Finally, for the T1 tumor class, the model got a precision of 0. 99, recall of 0. 99, and accuracy of 0%, precision of 0, and recall of 0, and the F1 score is 0. 99. The substantiation of the presented findings of 109 illustrations. These results prove the suitability of the ViT-BT when applied to differentiate the various kinds of brain tumors. Due to factors such as cost, time, and accuracy of the ViT-BT, it can positively impact the processes of medical diagnosis. Lastly, Table 4 compares the results obtained from ViT-BT with those of other pertinent studies, and it can be seen that our method has achieved better accuracy and efficiency. In another related work [43][52][53][54], the GAN (generative adversarial networks) model was trained on a 60% training dataset and was given an accuracy of 96. 25%. Another model derived from GAN [44-48] obtained 96% valid outcomes in the identification of tumors. Another is the BW-VGG19 in paper [49-51]. The CNN-based architecture was applied, and it performed an accuracy of 97% on the 70% training dataset. To achieve 97% accuracy, the MANet approach [42] was used. Machines were tested to have the ability to detect tumors in the brains at 71% accuracy. Model [49] experienced an accuracy of 98% for the participation. Our ViT-BT computed relatively faster than all the above papers and recognized 98. Of the complex types of brain tumors, the input correctly and reliably identified 13%. Applying more precise findings to the objectives would assist in improving tumor diagnosis and detection in the medical imaging sector.

Table 4. Comparison Results of different Methods

Paper	Method	Accuracy
[1]	Transformer + CNN	96.75%
[49]	Deep CNN	96%
[50]	SMO+SVM	93.9%
[51]	CNN+ SVM + kNN	97%
[43]	GAN	96.25%
[42]	MANet	97.7%
[49]	BW-VGG 19	98%
[20]	FN-ViT	98.13%
[46]	Encoder-skip connection-decoder (U-net)	98%
Our Method	ViT-BT	98.17%

As a result of this paper, the Vision Transformer model distinguished itself with the highest accuracy of 98. % in classifying tumor MRI images, this indicates that the self-attention mechanism used by Vision Transformers was influential in capturing outstanding features of images. However, it can be noted that its performance might be affected by the dataset type, the complexity of the task it is assigned to, and other parameters. When evaluating the models from the two datasets using multiple indexes and employing all the models, the latter would give a better picture of the model's ability

A. Computational Resources and Time Requirements

i. Computational Resources:

- Hardware Specifications: The effectiveness of deep learning models, including Vision Transformers, heavily relies on the underlying hardware. High-performance GPUs (Graphics Processing Units) or TPUs (Tensor Processing Units) are typically required to handle the large-scale computations involved in training these models. The type of GPUs used is NVIDIA Tesla V100, and the number of GPUs employed during training.

Retrieval Number: 100.1/ijsce.D364414040924
DOI: 10.35940/ijsce.D3644.14040924
Journal Website: www.ijsce.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Memory Requirements: Vision Transformers often require substantial memory resources, especially when dealing with high-resolution MRI images and large batch sizes. The GPU has 16 GB of RAM.

ii. *Training Time:*

Duration of Training: The time to train the ViT-BT model can vary significantly based on several factors, including the dataset size, model complexity, and the chosen hyperparameters (e.g., learning rate, batch size). The training process took 7 hours, and 50 epochs were used.

Optimization Techniques: Techniques such as transfer learning can reduce training time by starting with a pre-trained model. Mixed precision training can also speed up the training process and reduce memory usage.

iii. *Inference Time:*

- Speed of Predictions: The inference time (the time taken to make predictions on new MRI scans) is crucial for practical applications. The

- **Batch Processing**: The ability to process multiple images simultaneously (batch inference) can also be important. The inference time potentially ranges from 10-50 ms per image.

iv. *Practical Implications:*

- Integration into Clinical Workflows: Understanding the computational and time requirements is vital for integrating the ViT-BT model into clinical workflows. The model does not require excessive resources or time, so it is feasible for routine use in hospitals or clinics.

IV. DISCUSSION

As for the solution, the findings of this work tackled the efficiency of the proposed model that employs the Vision Transformer (ViT-BT) for the detection of the brain tumor, given its MRI images. The model resulted in a very high accuracy of 98. 17% on the dataset, accurately identifying the four types of brain tumors: T1 weight, T2 weight, T1 cerebral blood volume, and Flair. The confusion matrix is usually separated into four rows that define the tumors in the specified manner.

A. Limitations of the Proposed ViT-BT Model

i. Outcome of Rare Tumor Types
- Limited Training Data: The architecture of the ViT-BT model implies that its performance will depend on the training dataset, its quality, and its diversity. The model, in particular, performs relatively poorly when it is tested on standard cases that are rare or atypical, which have yet to be spotted or included in the benchmark data. One of the implications of such a limitation is that clinicians may fail to diagnose such rare diseases accurately or not diagnose them at all, only to find out much later that the symptoms were characteristic of the condition. Generalization Challenges: In this connection, it is worth admitting that the self-attention mechanism inherent in Vision Transformers is somewhat effective yet may only be suitable for some types of tumors, especially when they present themselves or their imaging features in a particular manner. Yet, owing to the low frequency of these rare types of tumors, the model may need to learn the distinguishing features that help in classification.

ii. Generalizability to Other Medical Imaging Tasks: Generalizability to Other Medical Imaging Tasks:
- Specificity of the Model: The ViT-BT model is created to enhance the classification of brain tumors only in MRI images. Its architecture and training method cannot be applied to other relevant medical image analysis tasks, including, yet not restricted to, the classification of lung nodules in CT or the detection of lesions in X-ray images. There are general trends in the information characteristics of each imaging modality and disease so that specific features may exist. Domain Adaptation: This means that the model's accuracy may depend on the specific medical imaging dataset or imaging institution, depending on factors such as the imaging protocol used, the type of scanner, and the cohort of patients included. This variability can pose problems when the model has to be used in similar contexts involving other patients without being retrained or fine-tuned, which is often impractical in clinical practice.

iii. Limited Training Data: This confirms that the performance of the ViT-BT model in an analogous manner profoundly depends on the training set's quality and richness. The clinician integrating the model into the operating room may end up found wanting when the tumor is rare and does not resemble the frequently used glioblastomas or meningiomas on which the model was trained. As a result, the performance of these machines may be low when diagnosing these rare conditions, which may imply that their diagnosis is missed or done inappropriately.

iv. Generalization Challenges: It is also important to notice that the self-attention mechanism of Vision Transformers can be quite strong but may not be as versatile foro all kinds of tumor presentations, most of which have rather distinct imagery and radiographic features. There might be a danger that during the training, the model will never see enough of these rare tumors to pick up all the features necessary for classification.

B. Statistical Analysis

Applying confidence intervals and p-values enhanced the validity of the results provided for the ViT-BT model. Confidence intervals give a range that the true value of a measure, such as accuracy, precision, and recall, is likely to be. This information provides reliable and stable model performance on differing datasets. We report a 95% accuracy interval, which means that if the model is tested on different samples from the same population, a range of accurate values may be obtained.

1. Enhancing the Validity of Results: Enhancing the Validity of Results:- Explaining the specifics of the experiments and the results achieved for the ViT-BT model, one should have incorporated the presented results' statistical confidence intervals and p-values. Confidence intervals give a range in which the actual value of a measure (e.g., accuracy, precision, recall) is likely to be a result and an idea about the stability of the model for different datasets.

When we read a statement that the model reported a 95% accuracy, this tells us the likely accuracy values one would get if the process is repeated on other samples from the same population.

2. Statistical Significance of Findings: The significance of the results can be established by the P-values obtained after the computation. Supplementary material from successful runs is provided to facilitate the comparison of ViT-BT with baselines or other more current algorithms for the given datasets, and p-values can be used to determine whether the observed differences were statistically significant or could practically be due to variance.

V. CONCLUSION

In conclusion, Vision Transformer is a solid deep-learning model proven to solve multiple tasks related to computer vision. Due to its long-range dependence and global property, it is more effective for image structures with complex features and small components, such as medical images. Thus, by performing Vision Transformer on the tumor brain MRI data (ViT-BT), one can develop rich advantages to detect or exclude the presence of brain tumors in medical images compared to other deep learning methods, including VGG16 and EfficientNet-B7. This will assist the clinicians and researchers in implementing accurate and early diagnosis of diseases, disease progression, and formation of appropriate treatments.

DECLARATION STATEMENT

I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been sponsored or funded by any organization or agency. The independence of this research is a crucial factor in affirming its impartiality, as it has been conducted without any external sway.
- **Ethical Approval and Consent to Participate:** The data provided in this article is exempt from the requirement for ethical approval or participant consent.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Authors Contributions:** The authorship of this article is contributed solely

REFERENCES

1. M. Al-ordain, A. Khan, Sliman A., et al. “Combining the transformer and convolution for efficient brain tumor classification using MRI images.” Applied Science, MDPI, 2023. https://doi.org/10.3390/app13063680
2. P., M.; Anbumani, G.; Theivendren, P.; Gopal, M. “An Overview of Brain Tumor. In Brain Tumors”, IntechOpen: London, UK, 2022.
3. L. H. Enjun Z. Long Chen at el. “A transformer-based generative adversarial network for brain tumor synthetics”. Front NeuroSci, 30 Nov. 2022, Volume 16, 2022 https://doi.org/10.3389/fnins.2022.1054948
4. Ostrom, Q.T.; Patil, N.; Cioffi, G.; Waite, K.; Kruchko, C.; Barnholtz-Sloan, J.S., “CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2013–2017”, Neuro Oncol. 2020. https://doi.org/10.1093/neuonc/noaa200
5. Park, J.; Y.G., “Brain Tumor Rehabilitation: Symptoms, Complications, and Treatment Strategy”, Brain Neurorehabilit. 2022. https://doi.org/10.12786/bn.2022.15.e25
6. Z.li, Y. Cong, Xiu Chen, et al. “Vision transformer-based weakly supervised histopathological image analysis of primary brain tumors,” Science volume 26, issue 1, 2023. https://doi.org/10.1016/j.isci.2022.105872
7. Bosman, F.T., “Integrative Molecular Tumor Classification: A Pathologist’s View”, In Encyclopedia of Cancer, 3rd ed.; Boffetta, P., Hainaut, P., Eds.; Academic Press: Oxford, UK, 2019; pp. 279–285.
8. Deng, J.; Hua, L.; Bian, L.; Chen, H.; Chen, L.; Cheng, H.; Dou, C.; Geng, D.; Hong, T.; Ji, H.; et al., “Molecular diagnosis and treatment of meningiomas: An expert consensus”, Chin. Med. J. 2022, 135, 1894–1912. https://doi.org/10.1097/CM9.0000000000002391
9. La Rosa, S.; Uccella, S., “Pituitary Tumors: Pathology and Genetics. In Encyclopedia of Cancer”, 3rd ed.; Boffetta, P., Hainaut, P., Eds.; Academic Press: Oxford, UK, 2019; pp. 241–256. https://doi.org/10.1016/B978-0-12-801238-3.65086-9
10. Zhang, L.; Liu, Y.; Huang, H.; Xie, H.; Zhang, B.; Xia, W.; Guo, B., “Multifunctional nano theranostics for near-infrared optical imaging-guided treatment of brain tumors,” Adv. Drug Deliv. Rev. 2022. https://doi.org/10.1016/j.addr.2022.114536
11. Brindle, K.M.; Izquierdo-García, J.; Lewis, D.; Mair, R.; Wright, A.J. Brain tumor imaging. J. Clin. Oncol. 2017. https://doi.org/10.1200/JCO.2017.72.7636
12. Jose B., Kaiser K., Daniel S., et al. “Deep convolutional neural networks for brain image analyses on magnetic resonance imaging: a review,” Volume 95, April, pages 64-81, Science Direct, Elsevier, 2019. https://doi.org/10.1016/j.artmed.2018.08.008
13. A. Arkinya, F. Zaccayn, James, et al. “Brain tumor diagnosis using medical learning, convolutional neural networks, capsule neural networks, and vision transformers, applied to MRI: a survey,” Journal of Imaging, volume 8, issue 8, 2022https://doi.org/10.3390/jimaging8080205
14. Kaiming He, Xiangyu Z., Shaoqing R., Jian S., “Deep residual learning for image recognition. Computer vision and pattern recognition”, doi.org/10.48550/arXiv.1512.03385, 2015.
15. Philip M., Harsh S. M., “Transfer learning with convolutional neural networks for classification of abdominal ultrasound images”, National Library of Medicine, 30 (12), 2017.
16. Bieza, A.; Krumina, G., “The value of magnetic resonance spectroscopy and diffusion tensor imaging in the characterization of gliomas growth patterns and treatment efficiency,” J. Biomed. Sci. Eng. 2013, 6, 518–526. 2017. https://doi.org/10.4236/jbise.2013.65066
17. Abd-Ellah, M.K.; Awad, A.I.; Khalaf, A.A.; Hamed, H.F., “A review on brain tumor diagnosis from MRI images: Practical implications, key achievements, and lessons learned,” Magn. Reson. Imaging 2019. https://doi.org/10.1016/j.mri.2019.05.028
18. Hemanth, G.; Janardhan, M.; Sujihelen, L., “Design and implementing brain tumor detection using machine learning approach”, In Proceedings of the 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 23–25 April 2019. https://doi.org/10.1109/ICOEI.2019.8862553
19. Salmon, E.; Ir, C.; Hustinx, R., “Pitfalls and Limitations of PET/CT in Brain Imaging,” Seminars in Nuclear Medicine; Elsevier: Amsterdam, The Netherlands, 2015; pp. 541–551. https://doi.org/10.1053/j.semnuclmed.2015.03.008
20. Abdullah A. Asiri L., Ahmad Shaf, Tariq Ali, et al.;” Exploring the Power of Deep Learning: Fine-Tuned Vision Transformer for Accurate and Efficient Brain Tumor Detection in MRI Scans”, MDPI, Diagnostics 2023, 13, 2094. https://doi.org/10.3390/diagnostics13122094
21. Swati, Z.N.K.; Zhao, Q.; Kabir, M.; Ali, F.; Ali, Z.; Ahmed, S.; Lu, J., “Brain tumor classification for MR images using transfer learning and fine-tuning”, Comput. Med. Imaging Graph 2019, 75, 34–46. https://doi.org/10.1016/j.compmedimag.2019.05.001
22. AlBadawy, E.A.; Saha, A.; Mazurowski, M.A., “Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing”, Med. Phys. 2018, 45, 1150–1158. https://doi.org/10.1002/mp.12752
23. Hongwei Bran Li, Gian Marco Conte, Syed Muhammad Anwar, et al.; “The Brain Tumor Segmentation (BraTS) Challenge 2023: Brain MR Image Synthesis for Tumor Segmentation (BraSyn)”, arXiv:2305.0901August 23 2023
24. Thaha, M.M.; Kumar, K.P.M.; Murugan, B.S.; Dhanasekeran, S.; Vijayakarthick, P.; Selvi, A.S., “Brain tumor segmentation using convolutional neural networks in MRI images”, J. Med. Syst. 2019, 43, 294. https://doi.org/10.1007/s10916-019-1416-0

25. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.-M.; Larochelle, H., "Brain tumor segmentation with deep neural networks," Med. Image Anal. 2017, 35, 18–31. https://doi.org/10.1016/j.media.2016.05.004
26. Sharif, M.I.; Li, J.; Amin, J.; Sharif, A., "An improved framework for brain tumor analysis using MRI based on YOLOv2 and convolutional neural network", Complex Intell. Syst. 2021, 7, 2023–2036. https://doi.org/10.1007/s40747-021-00310-3
27. Zein Eldin, R.A.; Karar, M.; Coburger, J.; Wirtz, C.; Burgert, O. DeepSeg: "Deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images", Int. J. Comput. Assist. Radiol. Surg. 2020, 15, 909–920. https://doi.org/10.1007/s11548-020-02186-z
28. Hatami Zadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Roth, H.; Xu, D., "Swin transformers for semantic segmentation of brain tumors in MRI images. In Brain lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries:", 7th International Workshop, 2021. https://doi.org/10.1007/978-3-031-08999-2_22
29. Jia, Q.; Shu, H. Bitr-Unet, "A CNN-transformer combined network for MRI brain tumor segmentation. In Brain lesion: Glioma, Multiple Sclerosis, Stroke, and Traumatic Brain Injuries", 7th International Workshop, Brain Les 2021. https://doi.org/10.1007/978-3-031-09002-8_1
30. Wang, W.; Chen, C.; Ding, M.; Yu, H.; Zha, S.; Li, J., "Trans bts: Multimodal brain tumor segmentation using transformer", in Proceedings of the Medical Image Computing and Computer, France, 27 September–1 October 2021. https://doi.org/10.1007/978-3-030-87193-2_11
31. Peiris, H.; Hayat, M.; Chen, Z.; Egan, G.; Harandi, M., "A robust volumetric transformer for accurate 3D tumor segmentation", In Proceedings of the Medical Image Computing and Computer Assisted Intervention—MICCAI 2022: 25th International Conference, Singapore, 18–22 September 2022. https://doi.org/10.1007/978-3-031-16443-9_16
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I., "Attention is all you need," in Advances in Neural Information Processing Systems: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017.
33. Sharma, K.; Kaur, A.; Gujral, S., "A review on various brain tumor detection techniques in brain MRI images", IOSR J. Eng. 2014, 4, 6–12. https://doi.org/10.9790/3021-04530612
34. Wiest, R.; Menze, B.; Reyes, M.; Porz, N.; Van Leemput, K., "The multimodal brain tumor image segmentation benchmark (BraTS)", IEEE Trans. Med. Imaging 2014, 34, 1993–2024.
35. Park, N.; Kim, S., "How do vision transformers work?" arXiv 2022, arXiv:220206709.
36. Özyurt, F.; Sert, E.; Avci, E.; Dogantekin, E.," Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy", Measurement 2019. https://doi.org/10.1016/j.measurement.2019.07.058
37. Polly, F.; Shil, S.; Hossain, M.; Ayman, A.; Jang, Y.M., "Detection and classification of HGG and LGG brain tumor using machine learning". In Proceedings of the 2018 International Conference on Information Networking (ICOIN), Chiang Mai, Thailand, 10–12 January 2018; pp. 813–817. https://doi.org/10.1109/ICOIN.2018.8343231
38. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al., "An image is worth 16 × 16 words: Transformers for image recognition at scale", arXiv 2020, arXiv201011929.
39. Steyaert, S.; Qiu, Y.; Zheng, Y.; Mukherjee, P.; Vogel, H.; Gevaert, O., "Multimodal data fusion of adult and pediatric brain tumors with deep learning". Med arXiv 2022. https://doi.org/10.1101/2022.09.21.22280223
40. Deepak, S.; Ameer, P. Brain tumor classification using deep CNN features via transfer learning. Computing. Biol. Med. 2019. https://doi.org/10.1016/j.compbiomed.2019.103345
41. Tummala, S.; Kadry, S.; Bukhari, S.; Rauf, H.T., "Classification of Brain Tumor from Magnetic Resonance Imaging Using Vision Transformers Ensembling". Curr. Oncol. 2022, 29, 7498–7511. https://doi.org/10.3390/curroncol29100590
42. Shaik, N.S.; Cherukuri, T.K., "Multi-level attention network: Application to brain tumor classification". Signal Image Video Process 2022, 16, 817–824. https://doi.org/10.1007/s11760-021-02022-0
43. Ahmad, B.; Sun, J.; You, Q.; Palade, V.; Mao, Z., "Brain tumor classification using a combination of variational autoencoders and generative adversarial networks". Biomedicines 2022, 10, 223. https://doi.org/10.3390/biomedicines10020223
44. Asiri, A.A.; Aamir, M.; Shaf, A.; Ali, T.; Zeeshan, M.; Irfan, M.; Alshamrani, K.A.; Alshamrani, H.A.; Alqahtani, F.F.; Alshehri, A.H.D., "Block-Wise Neural Network for Brain Tumor Identification in Magnetic Resonance Images". Comput. Mater. Contin. 2022, 73, 5735–5753. https://doi.org/10.32604/cmc.2022.031747
45. Asiri, A.A.; Shaf, A.; Ali, T.; Aamir, M.; Usman, A.; Irfan, M.; Al-Shamrani, H.A.; Mehdar, K.M.; Alshehri, O.M.; Alqhtani, S.M. "Multi-Level Deep Generative Adversarial Networks for Brain Tumor Classification on Magnetic Resonance Images". Intell. Autom. Soft Computing. 2023, 127–143. https://doi.org/10.32604/iasc.2023.032391
46. Asala Z., Khawla H., "Brain MRI Images segmentation based on U-Net architecture". Iraqi Journal for Electrical and Electronic Engineering (IJEEE). Volume 18, Issue 1, 2022. https://doi.org/10.37917/ijeee.18.1.3
47. Fabian I, Philip K., Wolfgang at el., "Brain tumor segmentation and radiomics survival predictions: Contribution to the BraTS2017 challenge", Lecture notes in computer science, pp 287-297, 2018. https://doi.org/10.1007/978-3-319-75238-9_25
48. Ayesh Younis, Li Qiang, Charles O., et al., "Brain tumor analysis using deep learning and VGG16 ensembling learning approaches". MDPI, Applied Science, 12(14), 2022. https://doi.org/10.3390/app12147282
49. Abu Bakr; Shadman Sakib; Mohammed M. et al., "Deep Convolutional Neural Networks Model Based Brain Tumor Detection in Brain MRI Images." Fourth International Conference on I-SMAC, India, 2020.
50. Heba Mohsen; El-Sayed E.; El-Syed M.; "Classification using deep learning neural networks for brain tumors". Future Computing and Informatics Journal, Vol. 3, Issue 1, Science Direct 2018. https://doi.org/10.1016/j.fcij.2017.12.001
51. M. O. Khairandish, M. Sharma, V. Jain, et al., "A Hybrid CNN-SVM Threshold Segmentation Approach for Tumor Detection and Classification of MRI Brain Images". IRBM, ELSEVIER, Vol. 43, Issue 4, 2022. https://doi.org/10.1016/j.irbm.2021.06.003
52. Ruria, S., Gautam, P., Raj, A., & Pandey, G. (2024). Brain Tumor Detection System using Deep Learning. In International Journal of Innovative Technology and Exploring Engineering (Vol. 13, Issue 3, pp. 23–27). https://doi.org/10.35940/ijitee.h9678.13030224
53. Agrawal, Y., & Birchha, V. (2020). Classification of Brain Tumour in MRI Images using BWT and SVM Classifier. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 6, pp. 3662–3667). https://doi.org/10.35940/ijrte.f7958.038620
54. Shetty, S., & Shetty, J. (2020). Classification of Brain Tumor using Convolutional Neural Networks. In International Journal of Engineering and Advanced Technology (Vol. 9, Issue 3, pp. 2841–2845). https://doi.org/10.35940/ijeat.c5995.029320

AUTHOR PROFILE

Khawla Hussein Ali has Assist. Prof. She has an M.Sc. in computer science from the College of Science, University of Basra, Iraq, and a Ph.D. from Huazhong University in China. She has been interested in and has published many papers on image processing, computer vision, signal processing, machine learning, deep learning, pattern recognition, and medical image processing. Her programming skills include C++, Python, Java, and Javascript. She has also taught coursework at the University on Computer vision, Pattern recognition, Operating systems, data structures, and data mining. Her Master's thesis is about image processing and enhancement of images using digital filters.