Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/jptcp.v31i7.7195

TRANSFORMING BRAIN TUMOR DIAGNOSIS: VISION TRANSFORMERS COMBINED WITH ENSEMBLE TECHNIQUES

Anees Tariq[1], Muhammad Munwar Iqbal[2*], Sumayya Bibi[3], Muhammad Hassan Butt[4], Shabana Ramzan[5]

[1,2*,4]Department of Computer Science, University of Engineering and Technology, Taxila
[3]Department of Communication Engineering, Faculty of Electrical Engineering, Universiti Teknologi, Malaysia 81310 UTM, Johar Bahru, Johar, Malaysia
[5]Department of Computer Science & IT, Govt Sadiq College Women University, Bahawalpur

***Corresponding Author:** Muhammad Munwar Iqbal
*Email: munwar.iq@uettaxila.edu.pk

Abstract
Brain Tumor (BT) is widely recognized as one of the most prevalent illnesses worldwide, affecting approximately 24,810 people in the year 2023. Most people suffering from brain tumor disease belong to the Southeast Asian and Western Pacific regions. Medical diagnostics using artificial intelligence and deep learning models demonstrate efficacy in addressing critical health challenges in initial disease and detection of intervention of BT. In this paper, we proposed ViT along with ensemble learning models for multiclass brain tumor classification and detection. The proposed work aims to provide the novel best solution to the problem of brain tumor detection using a deep learning approach. Ensemble Learning obtained 96% accuracy and loss of 0.13 with an F1-score, precision, and recall of 0.96. The comparative result shows that Vision Transformer ViT obtained 90% accuracy and loss of 0.30 with an F1-score, precision, and recall of 0.89 on the brain tumor MRI dataset containing 7023 images, which is further divided into train and test. The promising results showcase the potential of this proposed system in early and accurate brain tumor detection. The proposed system can be used in the early detection of brain tumors.

Keyword: Brain Tumor (BT), MRI dataset, Computerized Tomography (CT), Machine Learning, Deep Learning, Vision Transformers.

1. Introduction
Brain tumor identification is a significant part of medical image analysis for brain disease identification. Magnetic Resonance Imaging (MRI) and Computerized Tomography (CT) pictures are suitable for detecting brain malignancies [1]. AI systems, notably Deep Learning (DL)-based approaches, have made significant advances in brain tumor identification [2]. The suggested approach for brain tumor identification is separated into two parts: preprocessing and segmentation. Preprocessing uses local binary patterns, while segmentation uses techniques such as K-means, edge detection, and morphological operations like erosion and dilation [3]. Several machines and deep learning techniques have been developed for brain tumor diagnosis, including regression-based approaches that predict both the class and the bounding box coordinates in a single run and apply the full CNN to the entire picture to speed up the process [2]. Previous research has struggled with accurately recognizing tiny brain tumors in images to improve tumor detection accuracy. However,

the scientists sought to enhance this by gathering photos of brain tumors of variable sizes and utilizing a mix of position data from lower levels and detailed property data from upper layers to distinguish tumor pixels of different sizes [4]. Furthermore, CNN-based models paired with segmentation approaches were investigated and contrasted, with transfer-learning-based models such as ResNet50, Inceptionv3, and InceptionResNetv2 being utilized [1]. One study's authors developed a custom dataset with almost 1100 MRIs that included both Low-Grade-Gliomas (LGG) and High-Grade-Gliomas (HGG). They then suggested a two-pronged method for brain tumor detection that involved a small dataset's training using transfer learning and the YOLO detection framework [1, 2]. Furthermore, some researchers have proposed new approaches for brain tumor detection that use image compression strategies based on Deep Wavelet Auto-encoder (DWA), as well as spectrum angle-dependent feature extraction and Spectral Clustering Independent Component Analysis (SCICA) to improve brain tumor classification from MRI images [1].

Brain cancers are detected using a variety of approaches, including medical imaging. Deep learning models were used to detect brain cancers with great accuracy and precision in MRI scans. Comparison research was undertaken on several brain tumor detection techniques to determine their efficiency in detecting the existence of a tumor within the brain [3-5]. While computed tomography (CT) is less sensitive and specific, Magnetic Resonance Imaging (MRI) is more sensitive but less specific in detecting brain malignancies [1]. Texture analysis approaches using different classifiers have also been used to improve tumor identification in MRI images [6]. Convolutional Neural Networks (CNNs) have been used to accurately classify a variety of brain malignancies, including gliomas, meningiomas, and pituitary tumors [7]. Promising results have also been obtained when using fast object detection models, such as YOLOv2, YOLOv3, and Faster R-CNN, for brain tumor detection [2]. MRI image segmentation and tumor identification have also been studied using threshold-based, edge detection, region expanding, watershed, and K-means clustering algorithms [8]. Brain cancers are generally diagnosed by computed tomography scans, magnetic resonance imaging scans, and ultrasound imaging [9]. Numerous research gaps in brain tumor identification and diagnosis have been found by a systematic study of brain illnesses utilizing deep learning techniques [10].

The development of deep learning methods has made medical imaging a potentially useful tool for precisely identifying brain cancers [5]. For this reason, deep learning models have been put forth, and their performance in precisely locating and detecting brain tumors in MRI images has shown promise [4]. Nevertheless, the methods used today to detect brain tumors are still not without limits. CT is less sensitive than MRI, but it is more specific, according to a study comparing methods for detecting brain tumors [1]. In the meantime, a different study used three classifiers and four texture analysis techniques to evaluate and improve the detection of brain cancers [6]. Furthermore, researchers used CNN to classify various forms of brain malignancies from MRI images, including glioma, meningioma, and pituitary tumors [7]. Other research has looked into the use of quick object detection models for brain tumor detection, such as YOLO versions YOLOv2, YOLOv3, and RetinaNet [2]. In addition, numerous segmentation approaches are used for the study of brain tumor identification, including threshold-based, edge detection, region expanding, watershed, and k-mean clustering techniques [8]. The primary modalities for diagnosing brain tumors are CT, MRI, and ultrasound imaging [9]. Several upgraded machine-learning techniques have been created and compared in several research to improve the accuracy of brain tumor detection [11].

The major contribution of this research is to revolutionize brain tumor early detection and categorization by developing a robust computer-aided diagnosis tool. It would decrease the need for labor-consuming human treatments and increase the efficacy of treatment plans. Despite advances in deep learning and medical imaging approaches for brain tumor identification, further study is required to overcome the limitations of present systems.

The rest of the paper is organized as follows: Section II presents related work. Our proposed methodology is described in detail in Section III, and the results of our analysis are presented in Section IV. Finally, we conclude with a discussion of our findings and suggestions for future work in Section V.

2. Related Work:

Meshram et al. [12] focus on the use of data-driven approaches, notably deep learning models like ResNet 50 and Inception V3, to diagnose brain cancers using MRI images. The study provides a significant contribution by methodically curating matched datasets, each with unique properties. It incorporates critical strategies for hyperparameter optimization, such as Early Stopping and ReduceLROnPlateau. The implementation of strategic innovations, such as tailored pooling and regularization layers, results in high classification accuracy. In the dataset of four classes, ResNet 50 combined with the Nadam optimizer performed exceptionally well, with accuracy rates of 99.34% for gliomas, 93.52% for meningiomas, 98.68% for non-tumorous pictures, and 97.70% for pituitary tumors. ResNet50 with the Adam optimizer shines in a two-class dataset, achieving an overall accuracy of 97.84%, with perfect per-class accuracy of 97.62% for 'Tumor Positive' and 100% for 'Tumor Negative.' These findings highlight the revolutionary potential of the customized method, demonstrating significant advances in the area of brain tumor classification using MRI scans. Attention-based approaches have recently proved efficient in brain tumor categorization.

The ConvAttenMixer, described in [13], is a transformer model that combines convolutional layers, self-attention, and external attention processes to improve the processing of MRI brain pictures. By including two convolution mixer blocks, the model successfully captures spatial and channel-wise data relationships. Self-attention prioritizes local elements, whereas external attention focuses on global interactions. The model's classification head uses a squeeze-and-excitation mechanism. In performance evaluation, the ConvAttenMixer outperforms baseline models, reaching an excellent accuracy of 0.9794 on a dataset of 5712 MRI images. This accuracy beats baseline models, which range between 0.87 and 0.93. Notably, ConvAttenMixer has higher accuracy, recall, and f-measure, demonstrating its ability to analyze both local and global data with little computational memory needs.

Galic et al. [14] look at the enormous impact of artificial intelligence, namely deep learning, on the improvement of medical image processing and interpretation. The complete examination consists of a number of tasks, including sickness detection, categorization, and anatomical structure segmentation. The study highlights issues such as a lack of annotated data and image variability by objectively examining both strengths and limitations. It emphasizes the need for future research to concentrate on explainable deep learning approaches and the integration of multi-modal data.

Hussain et al. [15] focus on the critical global health issue of brain tumors, highlighting the importance of quick and precise diagnosis using MRI. It is the time-consuming nature of radiologists for manual evaluations. It is recognizing the limitations, there is a need for computer-aided categorization models. However, earlier models frequently encountered performance and explainability concerns, resulting in physician skepticism. To address this, the authors provide a novel classification and localization model that combines VGG-19, and EfficientNetV2 models with modified class activation mapping strategies. The experimental results reveal that the pre-trained VGG-19 with Grad Cam beats scratch VGG-19, EfficientNetV2, and other cutting-edge deep learning algorithms in terms of classification accuracy as well as visualization output. The proposed strategy has the potential to reduce diagnostic ambiguity while increasing the validity of brain tumor categorization.

The introduction of transfer learning in medical image analysis, particularly for 3D medical datasets, overcomes the issue of inadequate labeled data for training deep learning models. The study [16] offers a revolutionary technique known as the Medical Transformer. A multi-view technique used across three planes of the 3D volume improves spatial linkages while preserving parameter efficiency during training. The proposed methodology comprises pre-training the model with self-supervised learning on a large dataset of normal brain MRI images, with an emphasis on predicting masked encoding vectors. Assessment across three downstream objectives, detecting brain disorders, forecasting brain age, and segmenting brain tumors, demonstrates the Medical Transformer's improved performance over current transfer learning approaches. Notably, the technique reduces parameters by up to 92% for classification and regression tasks and 97% for segmentation, all while retaining strong performance even with a small number of training data.

As the frequency of brain tumors rises, researchers are increasingly focusing their efforts on automating the detection and diagnostic process using multi-tumor brain picture categorization. Deep neural networks have emerged as critical tools in medical picture categorization despite obstacles such as the vanishing gradient problem and overfitting. Recent research [17] offered a solution to overcome these concerns by adopting a deep network model that uses ResNet-50 and global average pooling. The suggested model was simulated using a dataset of 3064 pictures from three-tumor brain magnetic resonance imaging. The evaluation yielded remarkable results, with an average accuracy of 97.08% with data augmentation and 97.48% without data augmentation.

Luo et al. [18] present an approach to circumvent limits seen in typical Vision Transformer (ViT) models built for medical picture segmentation. The goal of this method is to improve generalizability across several organ areas. The proposed universal ViT segmentation model employs task-specific prompts to combine characteristics from the encoder of the ViT-based segmentation model and trainable universal prompts. This novel approach improves flexibility and efficiency by allowing for the segmentation of different organ areas inside a single model architecture. Empirical validation of several medical picture datasets confirms its efficacy and versatility. The results reveal increases in segmentation performance, demonstrating the promise of this forward-thinking technique for expanding medical image analysis and enabling a wide range of healthcare applications.

Since the introduction of MRI scans in 1978, researchers at EML Laboratories have opened the road for increased diagnostic capabilities. Primary malignant brain and CNS tumors were responsible for an expected 251,329 deaths in 2020. The traditional strategy of depending on histological subtyping during brain MRI interpretation creates subjectivity issues for radiologists. This problem is compounded in developing nations like India, where the doctor-to-population ratio is 1:1151. To overcome these issues, the study [12] uses swin transformers and deep learning approaches to identify, categorize, localize, and estimate the size of brain tumors in MRI images. The findings reveal a viable way to increase radiologists' and doctors' efficiency, providing a possible breakthrough in early tumor diagnosis and ultimately saving lives.

Ramakrishnan et al. [19] focus on the critical requirement for quick identification of malignant brain tumors and present a hybrid CNN architecture that incorporates InceptionV3, ResNet-50, VGG16, and DenseNet for classification. The approach employs a Kaggle dataset of 3929 pictures, including 2556 non-tumorigenic and 1373 tumorigenic cases, to extract features from MRI images, segment tumors using mask images, fuse segmented tumor images with originals, and classify them using four different CNN models. The models are further tuned using oneAPI to improve performance. The models are further tuned using oneAPI to improve performance. The evaluation uses mean Intersection over Union (mIoU) as a crucial statistic to determine the balance of genuine positives, false positives, and false negatives. This technique presents a viable option for rapid and accurate brain tumor classification, with optimization providing insights into individual model performance.

Table 1. Literature Review Summary

References	Year	Problem	Technique Used	Avg. Acc %	Limitation
S. Ahmmed et al. [12]	2023	Diagnose brain tumors based on magnetic resonance imaging images	ResNet 50, Inception V3	98%	Binary Class DataSet
S.M. Alzahrani et al. [13]	2023	Self-attention and external attention mechanisms to enhance the analysis of MRI brain images	ConvAttenMixer	88%	Small Dataset and Binary Class Dataset
K.Genereux et al. [14]	2023	Addressing the challenges associated with the complex anatomy of the brain and variability in hemorrhage appearance	2D CNN with Bi-LSTM	96%	Biased Dataset
E. Jun et al. [16]	2021	The emergence of transfer learning in medical image analysis, particularly for 3D medical datasets	Self-Supervised Learning	93%	Limited annotated data for training deep learning models

R. L. Kumar et al [17]	2021	A novel approach employing a deep network model utilizing ResNet-50 and global average pooling to mitigate overfitting problem	ResNet 50	96%	Small Dataset and No Data Augmentation
W. Luo et al. [18]	2023	Address limitations in conventional Vision Transformer (ViT) models for medical image segmentation.	Vision Transformers	77.3	Low accuracy
A. B. Ramakrishnan et al. [19]	2023	Focuses on the urgent need for rapid detection of malignant brain tumors	Hybrid CNN architecture	95%	Binary Class Dataset

3. Proposed Methodology

Figure 1 depicts our proposed framework's high-level design, which is comprised of three key components. The first module is data acquisition, which is a mechanism for gathering pictures of several types of brain tumors.

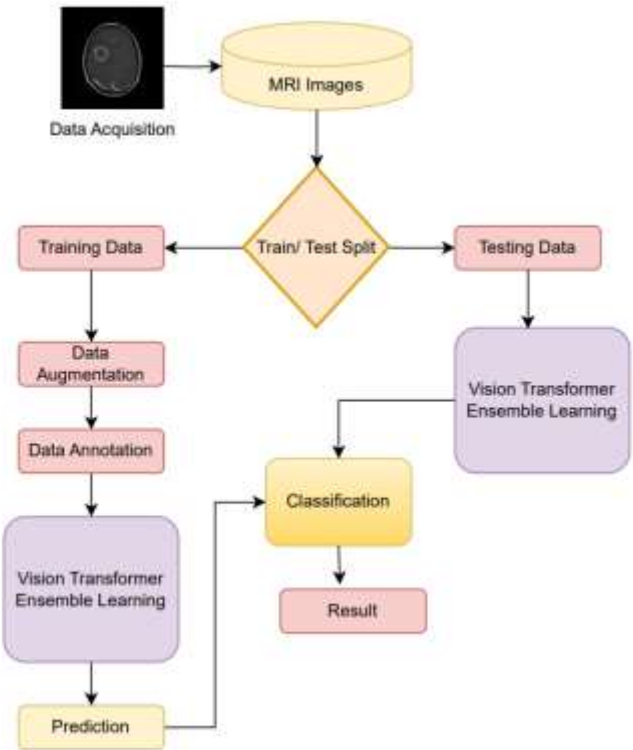

Figure 1. Block Diagram of Proposed Methodology

The second module, data preparation, is critical to our investigation. It involves data augmentation. The final module is classification, in which several deep-learning models are utilized to categorize brain cancers.

3.1. Data Acquisition

Any detection model's performance must be tested and assessed against a dataset. A standardized dataset is essential to get useful and effective results. We used the brain tumor MRI Dataset, which is an extensive dataset and contains brain tumor pictures from Kaggle. This dataset is freely available and contains almost 7,000 photos. Table 2. Represent dataset details. We incorporated the following dataset sets (figshare, SARTAJ dataset, and Br35H) into our dataset. This dataset includes 7023 photos of the human brain that are divided into four categories: glioma, meningioma, no tumor, and pituitary. Each class is separated into training and testing samples. The Glioma class has 1321 pictures for training and 300 for assessment. The Meningioma class includes 1339 photos for training and 306 for assessment. The No Tumor class has 1595 photos for training and 405 images for testing. The Pituitary class has 1457 pictures for training and 300 for assessment.

Table 2. Dataset Details

Class Label	Glioma	Meningioma	No tumor	pituitary	Total
Training Data Sample	1321	1339	1595	1457	5712
Testing Data Sample	300	306	405	300	1311
Total	1621	1645	2000	1757	7023

The entire database is randomly partitioned into training and test sets. We utilized 1311 photos for testing and 5712 images for training, which included various tumor types such as glioma, meningioma, no tumor, and pituitary. The 80/20 distribution is commonly employed in neural network applications; alternative distributions (such as 70/30 and 75/25) have insufficient influence on the performance of the produced model. This work employs 5712 photos to train the model and 1311 images to assess the model's performance for brain tumor classification.

3.2. Data Preprocessing

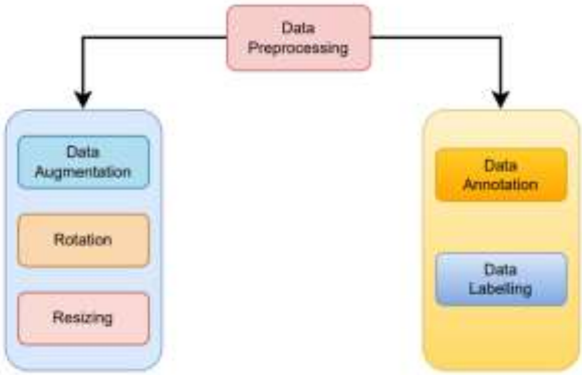

Figure 2. Data Preprocessing

Preprocessing is utilized to improve the performance of the proposed model by applying data preprocessing techniques (such as data augmentation) prior to brain tumor detection. We used data augmentation for data preparation. It is utilized to resize and rotate our images in various directions. We used data augmentation, as seen in Figure 3.

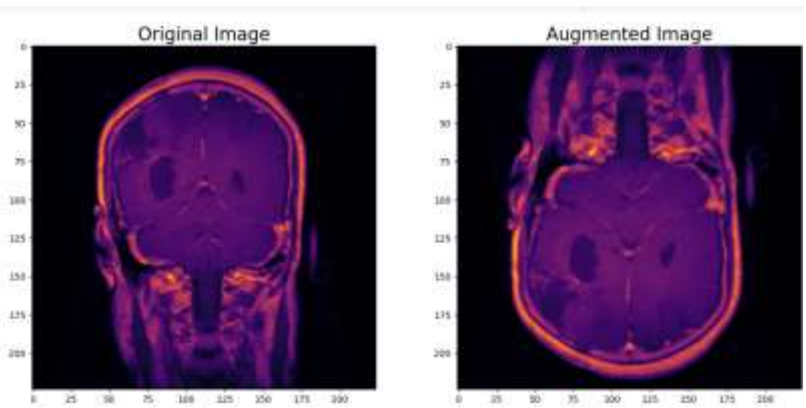

Figure 3. Data Augmentation

3.3. Vision Transformer

In ViT we input an image (I) which has some height, width, and number of channels defined as RHxWxC. Where "H" is height, "W" is width, "C" is the total number of channels, and R is the resolution of an image. An image is separated into N patches of different sizes, P x P x C, where N is $\frac{HW}{P^2}$. After that, position embeddings are added to these flattened picture patches to preserve the positioning information of the patches after linear embeddings have been computed for them. Figure 6 depicts the vision transformer model used for the MRI classification of multiclass brain tumors.

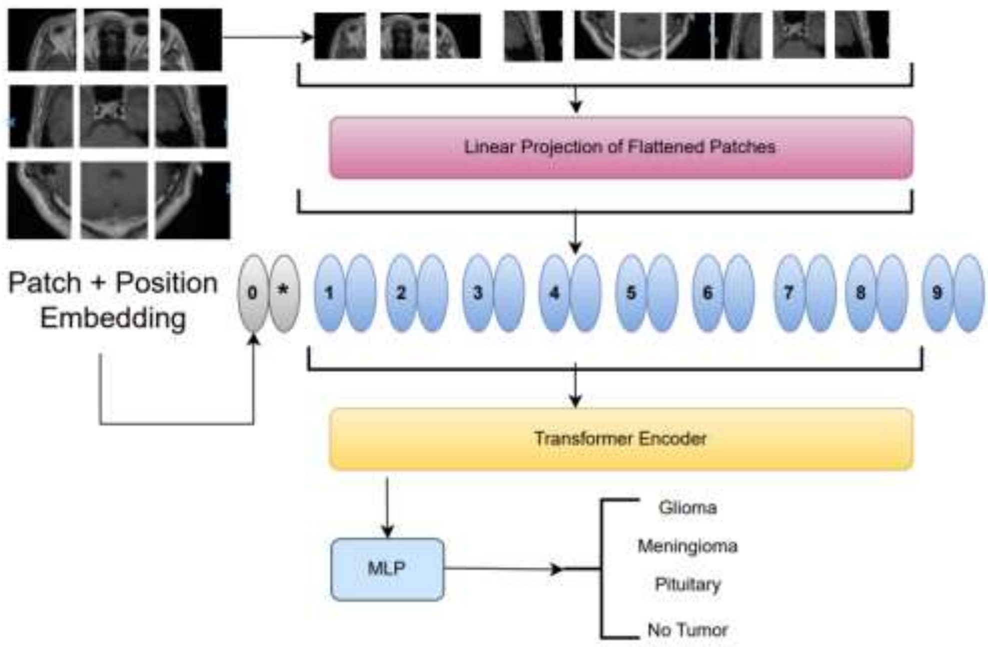

Figure 4. Vision Transformer Architecture.

A multilayer perceptron (MLP) head adds an extra learnable patch embedding for the final categorization. Furthermore, as illustrated in Figure 7. the transformer encoder model, which is made up of MLP blocks with layers of multi-headed self-attention, gets these combined position embeddings and patches.

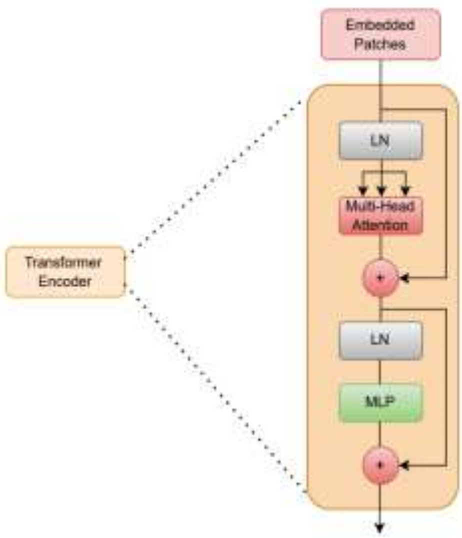

Figure 5. Encoder architecture for Vision Transformer.

In the Proposed work, a fine-tuned and pre-trained ViT b16 model was used where b stands for a base, and 16 indicates square patch size on the dataset. As a result, the MRI pictures were downsized to 224 by 224 resolution. The identical grayscale MRI picture is duplicated into the other two channels as the MRI slice only has one channel. Those ViT models that are pretrained require input into three channels.
In equation (4)-(7), ViT basic working is given.

$z_0 = [I_{class}; x_p^N E; \dots; x_p^N E] + E_{pos}$ E ∈ R(pxp,C)xD , Epos ∈ R(N+1)xD (1)

$Z'_l = MSA(LN(z_{l-1})) + z_{l-1}$ $l = 1 \dots L$ (2)

$z_l = MLP(LN(Z'_l)) + Z'_l$ $l = 1 \dots L$ (3)

$y = LN(Z_l^0)$ (7)

In equation (4), the positioning embedding was denoted by E_{pos}. The embedded path of N is denoted by x_p^N, and the output produced by the linear projection layer is denoted by z_0.
The transformer encoder layer's starting block is called layer normalization (LN). The next steps include residual connection and multi-head self-attention (MSA), resulting in an output. Z'_l at layer l. Equations (5) and (6) show how the second block begins with an LN layer, followed by an MLP and a residual link to output z_l. The transformer encoder model is seen in Figure 8. The transformer block's MLP is made up of two fully connected layers with GELU nonlinearity. After layer-normalization, equation (7) yields the final latent representation (with dimension D) of the input image I. During pretraining and fine-tuning, the final latent representation (Figure 8) is associated with the MLP or final classification heads.
Equation (8) describes the self-attention method. Where W, K, and B are the query, key, and value matrices created from matrix multiplication. The weights of the matrices $M_W, M_K, and\ M_B$ are learnable. It addresses the vanishing gradient problem, the product of the query, and the key is scaled with the square root of the size of the self-attention head as specified in equation (8).

$J = Attention(W, K, B) = softmax(\frac{WK^t}{\sqrt{D}})V$ (4)

$MSA(W, K, B) = [H_1, H_2, \dots H_h]M_0$ (5)

As illustrated in equation (9), where the total number of self-attention heads is denoted by H and M_0 is the learnable output transformation matrix, the concatenation of all self-attention heads is processed through a linear layer to create the MSA's final output.

3.4. Ensemble Learning

Combining the predictions of several base models into one ensemble approach called "ensembling via averaging" helps machine learning models become more stable and accurate. It minimizes the possibility of overfitting, and this method involves training each base model with distinct hyperparameters, methods, or feature subsets on the same dataset. It allows for the collection of various parts of the data. It uses an averaging approach, such as geometric mean, weighted average, or simple averaging. The predictions of the underlying models are integrated into a single forecast once they have been trained. Simple averaging is calculating the mean of each base model's predictions without taking into account the significance or performance of each model separately.
In contrast, weighted averaging gives each base model's forecast a weight determined by how well it performs on a validation set or other factors. Usually, methods like grid search and cross-validation are used to learn the weights. In situations involving positive-valued data, extreme values, or outliers, geometric mean averaging can be helpful in calculating the geometric mean of all base model predictions.
Model averaging ensembling, a class of ensemble techniques that combines the predictions of several models to increase forecast accuracy, includes weighted ensembling. Each model's prediction in weighted ensembling is given a weight that represents the model's relative performance or relevance. Many methods, such as grid search, cross-validation, or meta-learning, can be used to determine the weights allocated to each model's prediction. Weighted ensembling is a versatile and potent method

that can increase prediction accuracy and robustness by combining the advantages of several models and mitigating their disadvantages.
The equation (11) shows the weighted average ensembling.

$$\underline{y} = \frac{\sum_{j=1}^{n} w_j y_j}{\sum_{j=1}^{n} w_j} \quad (6)$$

Which expands to equation (7):

$$\underline{y} = \frac{w_1 y_1 + w_2 y_2 + \cdots + w_n y_n}{w_1 + w_2 + \cdots + w_n} \quad (7)$$

In machine learning, geometric mean ensembling is a sort of ensemble technique that combines the predictions of several base models by calculating the geometric mean of those predictions. It minimizes the possibility of overfitting, and this method involves training each base model with distinct hyperparameters, methods, or feature subsets on the same dataset. It allows for the collection of various parts of the data. The geometric means are used to merge the predictions of the basic models into a single forecast once they have been trained.
The geometric mean is computed by calculating the nth root of the product, where n is the number of models, and multiplying the forecasts of each base model together. The ensemble model's stability and robustness may be enhanced by the geometric mean, which tends to lessen the influence of extreme values or outliers on the predictions. For further accuracy and performance, it may also be used in conjunction with other ensemble approaches like bagging, weighted averaging, and simple averaging.
By reducing the variance and bias of the predictions, capturing a wider range of patterns and correlations in the data, and enhancing the overall accuracy and resilience of the model, geometric mean ensembling is a potent tool in ensemble learning.
Equation 13 explains the nth root of the product of n number.

$$\left(\prod_{j=1}^{n} y_j\right)^{\frac{1}{n}} = \sqrt[n]{y_1 y_2 \dots y_n} \quad (8)$$

4. Results and Discussions

The proposed system's performance is measured in the sub-metric of the confusion matrix. The sub-metric is accuracy, F1-score, recall, and precision. Specificity as accuracy rate determines the capacity to correctly classify brain tumors against the identified disease classes in this case. Precision is estimated by using Eq. (9), and the disease's capability for the model is computed using Eq. (10). The proposed approach did the accuracy classification, the F1-score is calculated, and also a weighted average representation of precision and recall. Eq. (11) is used to calculate the F1-score. The accuracy measure is another performance metric used for performance evaluation and is derived using Eq. (12). Eq. (13) is used to compute the error rate.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (9)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (10)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (11)$$

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Positive + Total\ Negative} \quad (12)$$

$$Error\ Rate = \frac{False\ Positive + False\ Negative}{Total\ Positive + Total\ Negative} \quad (13)$$

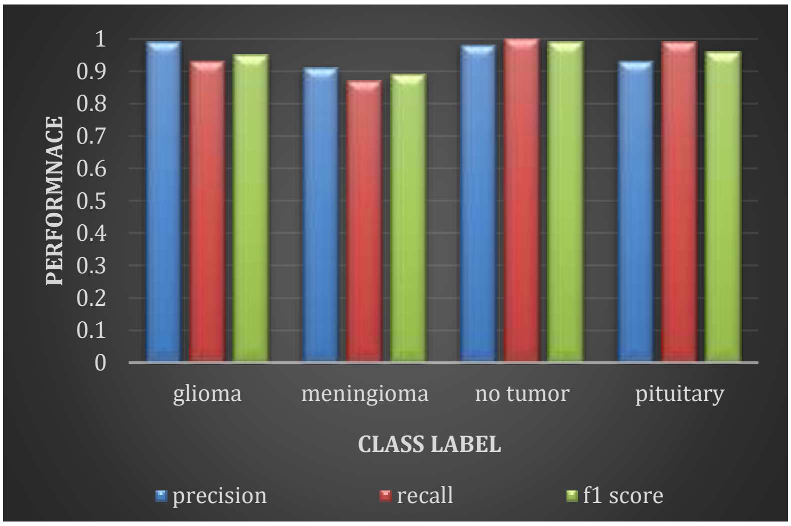

Figure 6. Precision, Recall, F1-score of Vision Transformers

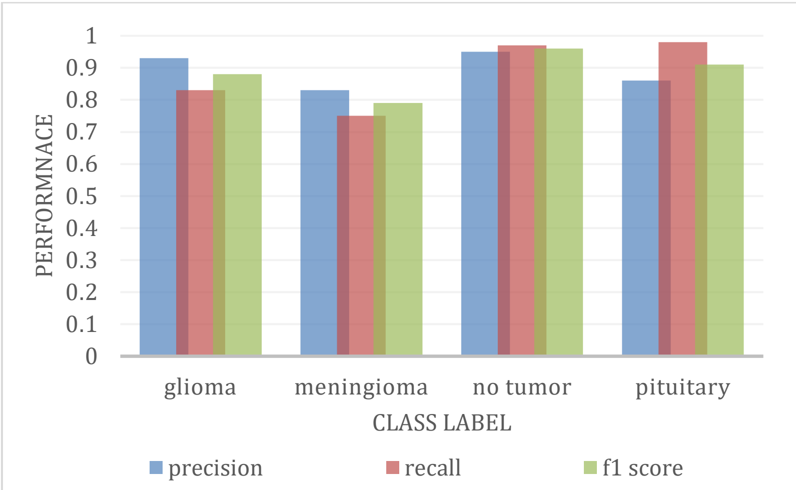

Figure 7. Precision, Recall, F1-score of Ensemble Learning

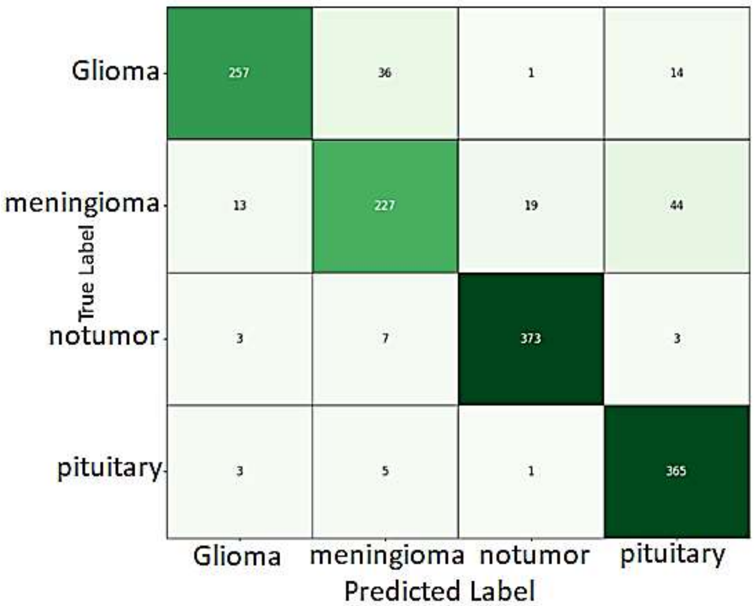

Figure 8. Confusion Matrix of Vision Transformer

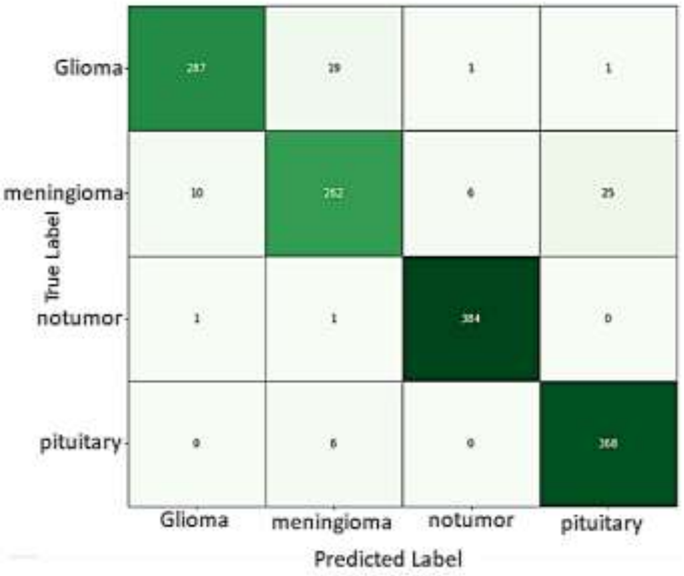

Figure 9. Confusion Matrix of Ensemble Learning

Fig. 10 displays two graphs, "Loss" and "Accuracy". The x-axis of both graphs is labeled "Epochs". The loss graph shows a general downward trend, while the accuracy graph shows a general upward trend as the number of epochs progresses. The training loss of the Vision Transformer is .30, and the validation loss is .34, while the accuracy of training and validation is 0.88 and 0.86.

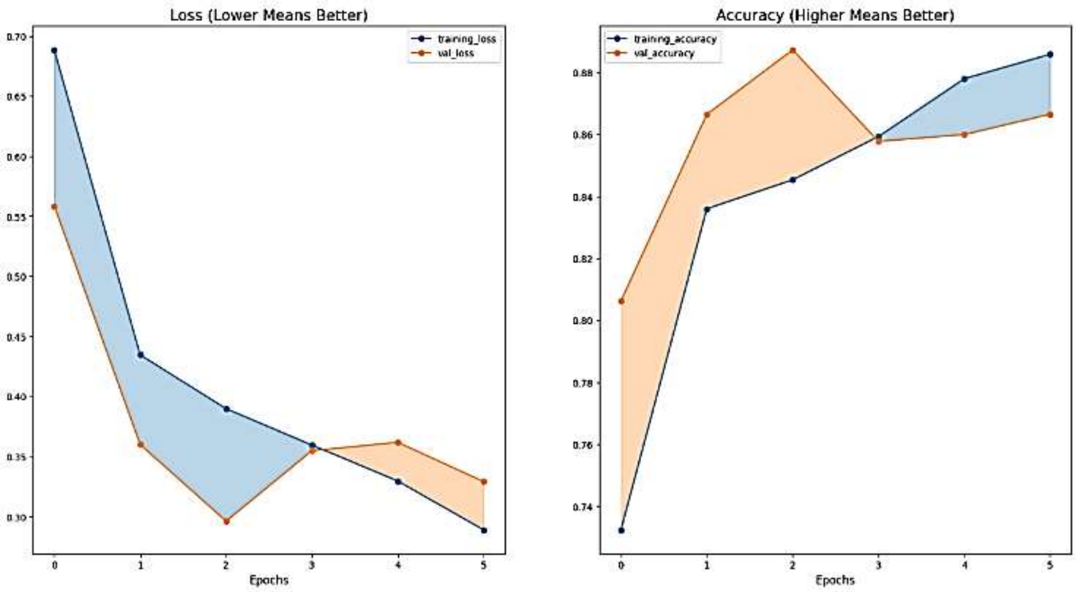

Figure 10. Model accuracy and loss graph of vision transformer

Table 3. Comparison of Proposed technique with existing techniques

References	**Year**	**Technique Used**	**Avg. Acc%**
E. Jun et al. [16]	021	Self-Supervised Learning	93%
K. Prathaban et al. [20]	023	MLS-CNN, ViT	91%
A. B. Ramakrishnan et al. [21]	023	Hybrid CNN architecture	95%
G. S. Tandel et al [22]	020	AlexNet	94%
Y. Zheng et al. [23]	023	UniVisNet	94%
Proposed Model	024	Vision Transformer, Ensemble Learning	89, 96%

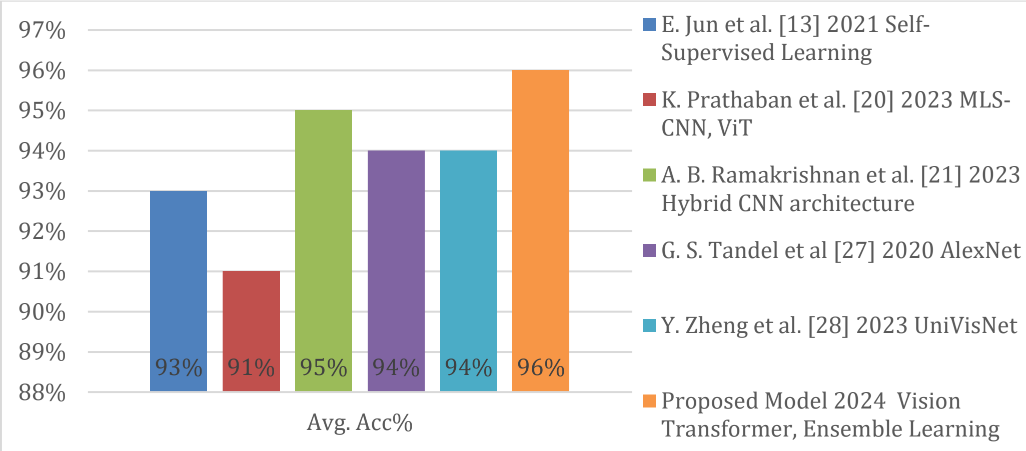

Figure 11. Comparison of Proposed Technique with Existing Techniques

5. Conclusions and Future Work

This study proposed multiclass brain tumor detection using an EfficientNetV2 and Vision Transformer. We performed our experiment on a publicly available dataset named the brain tumor MRI dataset. The proposed model successfully detects brain tumors using different images based on brain tumors. The sample comprised glioma, meningioma, no cancer, and pituitary class brain tumor. Ensemble Learning and ViT models were used, which are further given for average ensemble learning, weighted ensemble learning, and geometric mean ensemble learning. The proposed models give the highest accuracy of 96% with precision-recall and F1-score of 0.96 and a value loss of 0.13. Our experiments show that our proposed approach gives better results than existing approaches. We might build a new and improved deep learning model to improve efficiency in the future and provide a more accurate evaluation of the level of multiclass brain tumors. In the future, algorithms will be extended and capable of processing videos. It will be used a large and extended dataset with a variety of samples from international medical facilities to increase model resilience. The model's applicability and utility will be increased by addressing class imbalance through the development of synthetic data, thorough clinical validation, and adaptation to other imaging modalities like CT and PET scans.

References

1. Taher, F., et al., Efficient framework for brain tumor detection using different deep learning techniques. Frontiers in Public Health, 2022. **10**: p. 959667.
2. Hammami, M., S.K. Jarraya, and H. Ben-Abdallah, A comparative study of proposed moving object detection methods. Journal of Next Generation Information Technology. Volume, 2011. **2**(2).
3. Haritha, D. "Comparative study on brain tumor detection techniques", in 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES). 2016. IEEE.
4. Abdusalomov, A.B., M. Mukhiddinov, and T.K. Whangbo, Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 2023. **15**(16): p. 4172.
5. Gore, D.V., and V. Deshpande. Comparative study of various techniques using deep Learning for brain tumor detection, in 2020 International Conference for Emerging Technology (INCET). 2020. IEEE.
6. Bhagat, P., P. Choudhary, and K.M. Singh, A comparative study for brain tumor detection in MRI images using texture features, in Sensors for health monitoring. 2019, Elsevier. p. 259-287.
7. Khaliki, M.Z. and M.S. Başarslan, Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. Scientific Reports, 2024. **14**(1): p. 2664.
8. Pooja, V., M.K. Kumar, and K. Kamalesh, Comparative analysis of segmentation techniques on MRI brain tumor images. Materials Today: Proceedings, 2021. **47**: p. 109-114.

9. Akter, A., et al., Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor. Expert Systems with Applications, 2024. **238**: p. 122347.
10. Thapa, A., A. Roy, and S. Chakraborty, A comparative study of various metamodeling approaches in tunnel reliability analysis. Probabilistic Engineering Mechanics, 2024. **75**: p. 103553.
11. Habib, A.R.R., A comparative study of the machine learning-based energy management system for hydrogen fuel cell electric vehicles. Future Technology, 2024. **3**(1): p. 13-24.
12. Meshram, P.A., S. Joshi, and D. Mahajan, Brain Tumor Detection using Swin Transformers. arXiv preprint arXiv:2305.06025, 2023.
13. Alzahrani, S.M., ConvAttenMixer: Brain tumor detection and type classification using convolutional mixer with external and self-attention mechanisms. Journal of King Saud University-Computer and Information Sciences, 2023. **35**(10): p. 101810, 1319-1578.
14. Galić, I., et al., Machine Learning Empowering Personalized Medicine: A Comprehensive Review of Medical Image Analysis Methods. Electronics, 2023. **12**(21): p. 4411.
15. Hussain, T. and H. Shouno, Explainable Deep Learning Approach for Multiclass Brain Magnetic Resonance Imaging Tumor Classification and Localization Using Gradient-Weighted Class Activation Mapping. Information, 2023. **14**(12): p. 642.
16. Jun, E., et al., Medical transformer: Universal brain encoder for 3D MRI analysis. arXiv preprint arXiv:2104.13633, 2021.
17. Kumar, R.L., et al., Multiclass brain tumor classification using residual network and global average pooling. Multimedia Tools and Applications, 2021. **80**(9): p. 13429-13438.
18. Luo, W., et al. Universal Medical Image Segmentation with Task-Specific Prompt-Guided Transformer Model. 2023. IEEE.
19. Ramakrishnan, A.B., et al., Optimizing brain tumor classification with hybrid CNN architecture: Balancing accuracy and efficiency through oneAPI optimization. Informatics in Medicine Unlocked, 2024. **44**: p. 101436, 2352-9148.
20. Prathaban, K., et al., Detecting Tumor Infiltration in Diffuse Gliomas with Deep Learning. Advanced Intelligent Systems, 2023: p. 2300397.
21. Ramakrishnan, A.B., et al., Optimizing brain tumor classification with hybrid CNN architecture: Balancing accuracy and efficiency through oneAPI optimization. Informatics in Medicine Unlocked, 2023: p. 101436.
22. Tandel, G.S., et al., Multiclass magnetic resonance imaging brain tumor classification using artificial intelligence paradigm. Computers in Biology and Medicine, 2020. **122**: p. 103804.
23. Zheng, Y., et al., UniVisNet: A Unified Visualization and Classification Network for accurate grading of gliomas from MRI. Computers in Biology and Medicine, 2023. **165**: p. 107332.