JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Custom Vision Transformer Approach for Multiclass Brain Tumour Classification Using MRI Images

Asst. Prof. Dr. Sanju S. Anand[1] & Asst. Prof. Ashwini Dsouza[2]

Research Guides[1,2]; Student Researchers – Vaishnavi Naik, Varna Poojary, Shreya P., and Neeksha

Department of Computer Science, Alva's College (Autonomous), Moodubidri, Mangalore, Karnataka, India

Email: sanjualvascollege@gmail.com

ABSTRACT

Brain tumors are serious abnormalities that require careful medical evaluation. Magnetic Resonance Imaging (MRI) is widely used for examining brain structures and abnormalities, but manual interpretation can be time-consuming and requires specialized expertise. This study presents an automated image-classification approach using a Custom Vision Transformer (ViT) for brain MRI images.The proposed system classifies MRI images into four categories: glioma, meningioma, pituitary tumor, and no tumor. The project dataset contains 5,600 training images and 1,600 testing images. Images are converted to RGB format, resized to 224 × 224 pixels, converted to numerical arrays, and normalized before being supplied to the model.

The Custom Vision Transformer uses custom Patches and PatchEncoder layers. The trained model achieved a test accuracy of 86.81% and a test loss of 0.8091. The trained model was saved in HDF5 (.h5) format and integrated into a Streamlit application.The application provides a login interface, MRI image upload, predicted tumor category, confidence score, visual tumor/no-tumor indication, and basic information about the predicted category. The system is intended as an academic and research prototype and should not be used as a substitute for professional medical diagnosis.

Keywords

Brain Tumor, MRI, Vision Transformer, Deep Learning, Image Classification, TensorFlow, Keras, Streamlit.

1. INTRODUCTION

1.1 Background

Brain tumors are abnormal growths of cells in or around the brain. Depending on their type, location, and characteristics, they may affect neurological functions and require timely medical evaluation. MRI is an important imaging modality for examining brain abnormalities because it provides detailed information about soft tissues and brain structures.

Artificial Intelligence, Machine Learning, and Deep Learning have created new possibilities for computer-assisted medical image analysis. CNNs have traditionally been widely used because they can learn spatial image features automatically. More recently, Transformer-based architectures have attracted attention in computer vision because self-attention can model relationships between different regions of an input.

1.2 Problem Statement

Manual examination of large numbers of MRI images can be time-consuming. Automated classification is challenging because MRI images may vary in appearance, quality, and tumor characteristics. This project addresses development of a system that classifies brain MRI images into glioma, meningioma, pituitary tumour, and no tumour.

1.3 Motivation

The project is motivated by the opportunity to study Vision Transformer technology in medical image classification and connect model development with practical deployment through a Streamlit interface.

1.4 Objectives

- Develop a deep learning system for brain MRI classification.
- Implement a Custom Vision Transformer for four-class classification.
- Preprocess MRI images using a 224 × 224 input format.
- Evaluate the trained model on a separate testing dataset.
- Save the trained model in HDF5 (.h5) format.
- Integrate the model into a Streamlit web application.
- Display the predicted category and confidence score.

1.5 Significance and Scope

The study demonstrates an end-to-end workflow from image preprocessing and model development to deployment. The scope is limited to four-class image classification and does not include tumor segmentation, localization, treatment recommendation, or clinical diagnosis.

2. LITERATURE REVIEW

Artificial intelligence has been increasingly applied to brain tumor analysis using Magnetic Resonance Imaging (MRI). Earlier approaches primarily focused on conventional machine learning and Convolutional Neural Networks (CNNs), while more recent studies have explored Capsule Networks and Vision Transformers (ViTs). Akinyelu et al. reviewed machine learning, CNN, Capsule Network, and Vision Transformer approaches for brain tumor classification and segmentation. The survey discussed the advantages, limitations, open challenges, and future research directions associated with these approaches [1].

Deep learning-based CNN methods have demonstrated strong performance in brain tumor classification. Fabbri et al. investigated fully connected and convolutional neural networks for classifying glioma, meningioma, and pituitary tumors using T1-weighted contrast-enhanced MRI images. Their experiments achieved an average five-fold cross-validation accuracy of 91.43% for the best-performing neural network, demonstrating the potential of deep learning for automated brain tumor classification [2].

Transfer learning has also been widely investigated for MRI-based brain tumor classification. A study on deep CNN features using transfer learning employed a pretrained Google Net model for three-class classification of glioma, meningioma, and pituitary tumors. The proposed approach achieved a mean classification accuracy of 98% using patient-level five-fold cross-validation, demonstrating the effectiveness of pretrained CNN features for this task [3].

A systematic review of CNN-based brain tumor classification methods examined 83 studies published between 2015 and 2022. The review found that CNN architectures have been extensively applied to MRI-based brain tumor classification, although reported performance varied substantially depending on the dataset, classification problem, validation strategy, and evaluation metrics. The authors also highlighted the limited clinical usability of many existing studies [4].

Another study proposed an automated CNN-based approach for classifying glioma, meningioma, and pituitary tumors and compared the proposed model with pretrained architectures including VGG16, VGG19, ResNet50, MobileNetV2, and InceptionV3. The proposed approach achieved an accuracy of 98.04%, demonstrating the continuing effectiveness of CNN-based methods for brain tumor classification [5].

More recently, Vision Transformers have attracted attention because of their ability to model relationships between image regions using self-attention. Tummala et al. investigated pretrained and fine-tuned Vision Transformer models for three-class brain tumor classification using T1-weighted contrast-enhanced MRI images. Their dataset contained 3,064 MRI slices belonging to glioma, meningioma, and pituitary tumor classes. An ensemble of four ViT models achieved an overall testing accuracy of 98.7% at a resolution of 384 × 384 pixels [6].

Asiri et al. investigated fine-tuned pretrained Vision Transformer models for four-class brain tumor classification. Their study used 4,855 training images and 857 testing images. Several pretrained ViT architectures were evaluated, with ViT-B/32 achieving an accuracy of 98.24% [7].

Reddy et al. evaluated fine-tuned Vision Transformer models for multiclass brain tumor classification using 7,023 MRI images belonging to glioma, meningioma, pituitary, and no-tumor classes. The study compared the Transformer models with ResNet-50, MobileNet-V2, and EfficientNet-B0. Among the evaluated FTVT models, FTVT-L/16 achieved the highest accuracy of 98.70% [8].

Vision Transformers have also been investigated for brain tumor segmentation. A review of Transformer-based methods for multimodal brain tumor MRI segmentation discussed the use of pure and hybrid Transformer architectures and highlighted their ability to model non-local relationships in medical images. The review also discussed commonly used datasets, existing challenges, and future research directions for Transformer-based brain tumor analysis [9].

Krishnan et al. proposed a Rotation Invariant Vision Transformer for brain tumor classification from MRI images. The architecture incorporated rotation-related information into the Transformer representation to improve robustness to variations in image orientation. Their work demonstrates the development of specialized Transformer architectures for medical image analysis [10].

CNN-based methods and Transformer-based methods therefore represent two important directions in automated brain tumor classification. CNNs have demonstrated strong local feature extraction capabilities, while Vision Transformers provide an alternative approach based on image patches and self-attention(Table 1). The increasing use of Transformer architectures has created opportunities for developing models that can capture relationships between different regions of an MRI image.

Table 1. Summary of Selected Literature

Study	Approach	Dataset / Classes	Reported Result
Akinyelu et al. [1]	Survey	ML, CNN, CapsNet, ViT	Review
Fabbri et al. [2]	CNN / neural networks	3 classes	91.43%
Deep CNN transfer learning [3]	GoogLeNet transfer learning	3 classes	98%
CNN systematic review [4]	CNN-based methods	83 studies	91.63–100% for selected 3-class studies
Automated CNN study [5]	CNN	3 classes	98.04%
Tummala et al. [6]	ViT ensemble	3 classes; 3,064 MRI	98.7%
Asiri et al. [7]	Fine-tuned ViT	4 classes; 4,855 + 857	98.24%
Reddy et al. [8]	Fine-tuned ViT	4 classes; 7,023	98.70%

Study	Approach	Dataset / Classes	Reported Result
Transformer segmentation review [9]	Transformer-based methods	Multimodal brain MRI	Review
Krishnan et al. [10]	Rotation-Invariant ViT	Brain MRI	RViT approach

2.1 Research Gap

Existing studies demonstrate the strong potential of deep learning and Vision Transformer approaches for automated brain tumor analysis. CNN-based methods have achieved high classification performance through convolutional feature extraction and transfer learning, while recent Vision Transformer studies have reported strong results using pretrained, fine-tuned, ensemble, and specialized Transformer architectures. However, the reported results are based on different datasets, image resolutions, architectures, preprocessing techniques, and validation procedures, making direct comparison difficult.

In addition, many existing studies focus primarily on model performance, while practical integration of a trained model into a simple user-facing application is less emphasized. Therefore, this project focuses on developing a Custom Vision Transformer for four-class brain MRI classification involving glioma, meningioma, pituitary tumor, and no tumor, together with a Streamlit-based application for image upload and prediction.

The proposed project is intended as an academic prototype and not as a replacement for professional medical diagnosis.

3. METHODOLOGY

3.1 Dataset

The project dataset contains four classes: glioma, meningioma, no tumor, and pituitary tumor. It contains 5,600 training images and 1,600 testing images, giving 7,200 images in total(Table 2).

Dataset	**Images**
Training	**5,600**
Testing	**1,600**
Total	**7,200**

Table 2. Dataset distribution

3.2 Image Preprocessing

Each image is converted to RGB format, resized to 224 × 224 pixels, converted to a numerical array, converted to floating-point values, and normalized by dividing pixel values by 255.0. A batch dimension is added before prediction.

3.3 Class Labels

- Glioma → 0
- Meningioma → 1
- No Tumor → 2
- Pituitary → 3

3.4 Patch Extraction

The custom Patches layer divides the input image into smaller patches using TensorFlow's extract_patches operation. The patches are reshaped into a sequence for Transformer processing.

3.5 Patch Encoding

The PatchEncoder layer applies a dense projection to extracted patches and adds positional embeddings, helping retain patch-location information.

3.6 Custom Vision Transformer

The encoded patches are processed using Transformer-based feature learning and the resulting representation is used for four-class classification.

3.7 Model Training

The model is trained using the prepared training dataset and evaluated using the separate testing dataset. During training, it learns visual patterns associated with the four target categories.

3.8 Model Evaluation

The final evaluation produced a test accuracy of 86.81% and a test loss of 0.8091.

Metric	Result
Test Accuracy	86.81%
Test Loss	0.8091

Table 3. Model evaluation

3.9 Model Saving and Loading

The trained model is saved as brain_tumor_vit.h5. The application loads this saved model using TensorFlow/Keras together with the custom Patches and PatchEncoder layers. The model therefore does not need to be retrained each time the application starts(Table-3).

3.10 Prediction Pipeline

The prediction pipeline is: uploaded MRI image → RGB conversion → 224 × 224 resizing → normalization → batch expansion → model prediction → highest-probability class → confidence display.

3.11 Deployment

The saved model is integrated into a Streamlit application, separating the computationally intensive training stage from the user-facing inference stage.

4. SYSTEM DESIGN AND IMPLEMENTATION

4.1 System Architecture

The system contains a deep learning component and a Streamlit application component. The deep learning component learns patterns from MRI images and produces four-class predictions. The application provides login, image upload, prediction display, confidence, and basic tumor information.

4.2 Login Interface

The application begins with a doctor login interface containing username and password fields and a login button. The interface uses a brain MRI background and a clear application title.

4.3 MRI Upload Interface

After login, the user reaches the main Brain Tumor Detection page. An MRI image can be selected through the uploader and displayed after upload. A logout option is also available.

4.4 Prediction Interface

The prediction screen presents the general tumor/no-tumor status, predicted category, and confidence percentage. A green indication is used for No Tumor and a red indication for tumor predictions.

4.5 No-Tumor Prediction Result

The no-tumor result screen displays a green status message, the predicted class, confidence value, and basic description and prevention/management information.

4.6 Technologies Used

Technology	Purpose
Python	Application/model code
TensorFlow	Deep learning
Keras	Model construction/loading
NumPy	Numerical processing
PIL	Image processing
Streamlit	Web interface
HDF5 (.h5)	Model storage

Table 4. Technologies used in the system

4.7 Safety Statement

The application is intended for educational and project demonstration purposes and should not replace professional medical diagnosis.

5. RESULTS AND DISCUSSION

5.1 Model Performance

The Custom Vision Transformer achieved 86.81% test accuracy and 0.8091 test loss on the testing dataset. The result demonstrates that the model learned useful patterns for the four-class task, while incorrect classifications remain possible.

5.2 Confusion Matrix

The confusion matrix represents the class-wise performance of the proposed Vision Transformer model on the test dataset. It shows the number of correctly and incorrectly classified MRI images for glioma, meningioma, no tumor, and pituitary tumor. The diagonal elements represent correctly classified images, while the off-diagonal elements represent misclassifications (figure 5.1).

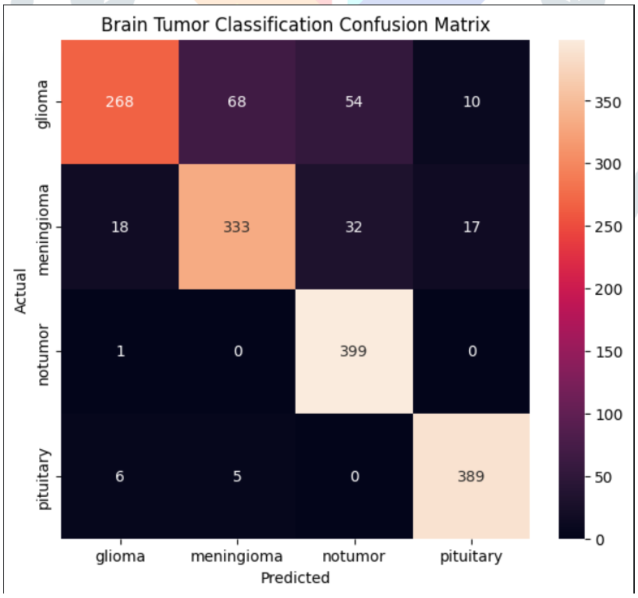

Figure 5.1.

Confusion Matrix of the Proposed Model

The confusion matrix shows that the model correctly classified 268 glioma images, 333 meningioma images, 399 no-tumor images, and 389 pituitary tumor images. The highest number of correct predictions was observed for the no-tumor class. Some misclassifications occurred between glioma and meningioma, indicating that these classes were comparatively more challenging for the model.

5.3 Classification Categories

- Glioma — Tumor detected
- Meningioma — Tumor detected

- Pituitary — Tumor detected
- No Tumor — No tumor detected

5.4 Confidence Score

The application calculates the confidence display from the maximum prediction probability. This indicates how strongly the model favors the selected class among the four outputs, but it should not be interpreted as clinical certainty.

5.5 Individual Prediction Examples

The demonstrated application results include a Pituitary classification with a displayed confidence of 99.93% and a No Tumor classification with a displayed confidence of 89.76%. These examples illustrate how the deployed system presents predictions to the user.

5.6 Prediction Behaviour

During individual testing, some tumor-containing images were occasionally predicted as No Tumor, particularly examples involving glioma and meningioma. This shows that overall test accuracy does not guarantee correct classification for every individual image.

5.7 Discussion

Several studies report higher accuracy using pretrained, fine-tuned, ensemble, or specialized Vision Transformer models. Direct numerical comparison is limited because datasets, resolutions, architectures, and procedures differ. The present project emphasizes a Custom ViT implementation and practical deployment.

5.8 Strengths

- Four-class MRI image classification.
- Custom patch extraction and positional encoding.
- Saved .h5 model for reuse without retraining.
- Interactive Streamlit interface.
- Green/red visual prediction indicators.
- Confidence score and basic class information.

5.9 Limitations

- Test accuracy is 86.81%, so incorrect classifications can occur.
- Performance depends on the dataset used for development.
- The system performs classification rather than tumor localization or segmentation.
- Confidence is a model output, not clinical certainty.
- The system has not undergone clinical validation.
- Images from other sources may produce different results.

5.10 Overall Findings

The developed system successfully connects a Custom Vision Transformer to a working Streamlit application. The model produces four-class predictions while the application presents the output in a user-friendly format. The observed errors show the need for further model improvement and validation.

5.11 Practical Significance

Separating model training from inference makes the system more practical because the computationally intensive training process does not have to be repeated whenever the application is launched.

6. CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

This study presented a Custom Vision Transformer-based system for brain MRI image classification. The system classifies MRI images into glioma, meningioma, pituitary tumor, and no tumor. Images are converted to RGB, resized to 224 × 224 pixels, converted to numerical arrays, and normalized before inference.

The Custom Vision Transformer uses Patches and PatchEncoder components to represent image regions as a sequence with positional information. The trained model achieved 86.81% test accuracy and 0.8091 test loss and was saved in HDF5 (.h5) format.

The model was integrated into a Streamlit application containing a login page, MRI upload interface, prediction result, confidence score, visual status indicators, and basic tumor information. The system demonstrates the practical deployment of a deep learning image-classification model.

Because the system can make incorrect predictions and has not undergone clinical validation, it should be considered an academic and research prototype rather than a clinical diagnostic tool.

6.2 Future Scope

- Use a larger and more diverse MRI dataset.
- Explore data augmentation and hyperparameter optimization.
- Compare Custom ViT with pretrained ViT, CNN, and hybrid architectures.
- Add tumor localization or segmentation.
- Add explainable AI methods to visualize important image regions.
- Evaluate using clinically validated datasets.
- Conduct expert and clinical validation before real-world use.
- Extend the application with reporting and improved visualization while maintaining privacy and security.

7. REFERENCES

[1] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, "Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey," *Journal of Imaging*, vol. 8, no. 8, p. 205, 2022, doi: 10.3390/jimaging8080205.

[2] J. S. Paul, A. J. Plassard, B. A. Landman, and D. Fabbri, "Deep Learning for Brain Tumor Classification," *Proceedings of SPIE—The International Society for Optical Engineering*, vol. 10137, 2017, doi: 10.1117/12.2254195.

[3] S. Deepak and P. M. Ameer, "Brain tumor classification using deep CNN features via transfer learning," *Computers in Biology and Medicine*, vol. 111, p. 103345, 2019, doi: 10.1016/j.compbiomed.2019.103345.

[4] P. Wang, Q. Yang, Z. He, and Y. Yuan, "Vision Transformers in Multi-modal Brain Tumor MRI Segmentation: A Review," *Meta-Radiology*, vol. 1, no. 1, p. 100004, 2023, doi: 10.1016/j.metrad.2023.100004.

[5] Z. Rasheed, Y.-K. Ma, I. Ullah, T. Al Shloul, A. B. Tufail, Y. Y. Ghadi, M. Z. Khan, and H. G. Mohamed, "Automated Classification of Brain Tumors from Magnetic Resonance Imaging Using Deep Learning," *Brain Sciences*, vol. 13, no. 4, p. 602, 2023, doi: 10.3390/brainsci13040602.

[6] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, "Classification of Brain Tumor from Magnetic Resonance Imaging Using Vision Transformers Ensembling," *Current Oncology*, vol. 29, no. 10, pp. 7498–7511, 2022, doi: 10.3390/curroncol29100590.

[7] A. A. Asiri, A. Shaf, T. Ali, M. A. Pasha, M. Aamir, M. Irfan, S. Alqahtani, A. J. Alghamdi, A. H. Alghamdi, A. F. A. Alshamrani, M. Alelyani, and S. Alamri, "Advancing Brain Tumor Classification through Fine-Tuned Vision Transformers: A Comparative Study of Pre-Trained Models," *Sensors*, vol. 23, no. 18, p. 7913, 2023, doi: 10.3390/s23187913.

[8] C. K. K. Reddy, P. A. Reddy, H. Janapati, B. Assiri, M. Shuaib, S. Alam, and A. Sheneamer, "A fine-tuned vision transformer based enhanced multi-class brain tumor classification using MRI scan imagery," *Frontiers in Oncology*, vol. 14, p. 1400341, 2024, doi: 10.3389/fonc.2024.1400341.

[9] P. T. Krishnan, P. Krishnadoss, M. Khandelwal, D. Gupta, A. Nihaal, and T. S. Kumar, “Enhancing brain tumor detection in MRI with a rotation invariant Vision Transformer,” *Frontiers in Neuroinformatics*, vol. 18, p. 1414925, 2024, doi: 10.3389/fninf.2024.1414925.

[10] H. Rostampour et al., “Deep learning approaches for automated classification and segmentation of head and neck cancers and brain tumors in magnetic resonance images: a meta-analysis study,” *European Journal of Radiology*, 2021. The study reviewed 30 eligible studies and reported pooled performance for deep-learning-based brain-tumor classification and segmentation.

****