© December 2025| IJIRT | Volume 12 Issue 7 | ISSN: 2349-6002

# Brain Tumor Detection Using Hugging Face

Dattatray Takale[1], Shravni Mastud[2], Sarthak Markad[3], Om Matere[4], Nayan Nagendra[5]
[1,2,3,4,5]*Department of Engineering, Science and Humanities (DESH)*
*Vishwakarma Institute of Technology, Pune, Maharashtra, India*

*Abstract* — **Brain tumors demand early, precise diagnosis for optimal treatment outcomes. This study presents an AI-based brain tumor detection system using Hugging Face's Vision Transformer (ViT) model integrated with a web frontend built in HTML, CSS, and JavaScript for user-friendly MRI image classification. The system enables users to upload MRI scans via a responsive web interface, processes them through the fine-tuned ViT model on Google Colab backend, and displays tumor classifications (glioma, meningioma, pituitary, no tumor) with confidence scores. Experimental results show 98-99% accuracy, validating its effectiveness for accessible medical screening.**

**Index Terms— Brain MRI Classification, Deep Learning, Hugging Face, Vision Transformer, Web-Based Diagnosis.**

## I. INTRODUCTION

Brain tumours, characterised by abnormal cell growth within the brain, pose significant risks to neurological function and patient survival. Early detection and precise classification into tumor subtypes such as glioma, meningioma, and pituitary are crucial for devising appropriate therapeutic interventions and improving prognosis. However, manual inspection of Magnetic Resonance Imaging (MRI) scans is plagued by lengthy analysis times, subjective interpretations, and limited accessibility in under-resourced areas. Addressing these challenges, this research utilises the cutting-edge Vision Transformer (ViT) model architecture, renowned for its superior performance in image classification tasks, particularly medical imaging. The ViT model is fine-tuned on extensive, annotated brain MRI datasets to accurately distinguish among tumor classes. Complementing the AI backend is a meticulously designed web frontend built with HTML for structure, CSS for styling, and JavaScript for interactivity, allowing users to easily upload MRI images, trigger predictions, and visualise results in an interactive, responsive environment. Deploying the model on Google Colab's GPU-enabled infrastructure with a RESTful API connection, the system bridges powerful AI with web accessibility, democratising advanced brain tumor screening tools for clinical and remote applications.

## II. LITERATURE REVIEW

Patel and Desai [1] proposed a hybrid approach combining Convolutional Neural Networks (CNN) with Support Vector Machines (SVM) for MRI-based brain tumor detection, demonstrating improved classification accuracy compared to standalone deep learning models. Saxena and Sharma [2] focused on deep learning-based MRI classification using a CNN architecture, showing that deep learning frameworks outperform traditional image-processing methods in tumor identification and grading. Khan and Hussain [3] explored transfer learning models for multi-class tumor classification, concluding that pretrained architectures such as ResNet-50 and VGG-19 significantly enhance accuracy and reduce training time due to their ability to generalize from large-scale medical imaging datasets .Mohammed and Yadav [4] investigated CNN-based segmentation techniques, emphasizing that automated segmentation enables precise boundary identification, which is essential for surgical planning and early diagnosis. Reddy and Banerjee [5] compared VGG-16, ResNet-50, and MobileNet for MRI tumor classification, reporting that lightweight models like MobileNet provide faster inference with minimal performance loss, making them suitable for real-time medical applications. Singh and Thomas [6] presented a lightweight deep learning framework capable of real-time tumor classification with reduced computational requirements, supporting deployment in low-resource clinical environments. Finally, Zhang and Liu [7] applied traditional machine learning and image-processing techniques for initial tumor localization and feature extraction, demonstrating foundational contributions that paved the way for advanced deep learning-based detection systems.

## III. METHODOLOGY

The system is developed using a fine-tuned Vision Transformer model (ViT-base-patch16-224) sourced from Hugging Face. The model is trained on a large brain MRI dataset containing more than 75,000 labelled images categorised into glioma, meningioma, pituitary tumor, and no-tumor classes. Before training, the dataset is cleaned to remove corrupted or mislabeled samples, and stratified sampling is used to split the data into training, validation, and testing sets while maintaining balanced class distribution.

All MRI images are resized to 224 × 224 resolution, converted to RGB if necessary, and normalised based on ImageNet statistics to ensure compatibility with the pretrained ViT weights. To improve model robustness and reduce the likelihood of overfitting, several augmentation techniques are applied, including random rotations, flips, translations, and controlled brightness or intensity changes. The original classification head of the ViT model is replaced with a custom fully connected layer adapted to the four target tumor classes. Dropout and batch normalisation layers are also added to improve stability and generalisation.

During fine-tuning, earlier transformer layers are either frozen or trained with a lower learning rate, while the newly added classification head is trained at a higher learning rate using the AdamW optimiser combined with a cosine or step-based learning rate scheduler. Categorical cross-entropy is used as the loss function to optimise multi-class classification. Training is performed on a GPU-enabled Google Colab environment using mini-batch gradient descent. Hyperparameters such as batch size, learning rate, and number of epochs are tuned through experimentation. After each training epoch, validation metrics, including accuracy, precision, recall, F1-score, and confusion matrices, are generated to monitor progress and detect misclassification patterns, such as similarities between glioma and meningioma. Early stopping and model checkpointing are used to store only the best-performing model weights.

To ensure reliability and generalisation, k-fold cross-validation is conducted, and performance results across folds are averaged. The final model achieves state-of-the-art accuracy above 98% with similarly high F1-scores, outperforming multiple conventional CNN-based approaches. Additionally, explainability techniques such as Grad-CAM and transformer attention visualisation are used to verify that the model focuses on relevant tumor regions, improving interpretability and medical trust.

The user interface for the system is developed using HTML5, CSS, and Bootstrap to ensure responsive behaviour across different screen sizes. JavaScript is responsible for handling tasks such as drag-and-drop image upload, real-time image previews using the Canvas API, and asynchronous communication with the backend using the Fetch API. Predictions returned in JSON format include the identified tumor type and associated confidence percentages, which are visually presented using interactive elements such as probability bars, meters, or charts. To improve usability, the interface includes loading indicators, disabled buttons during processing, and error messages for unsupported file types.

The trained ViT model is deployed as a REST API hosted on a GPU-powered Google Colab environment using ngrok for secure remote access. The backend receives incoming base64-encoded MRI scans, applies the same preprocessing pipeline used during training, forwards the image to the ViT model, and returns prediction results in structured JSON format. The system responds within approximately one second per image, enabling real-time diagnostic support.

Extensive testing is performed to validate functionality, reliability, and user experience. Cross-browser and device-level testing confirm compatibility with major desktop and mobile platforms. Functional testing verifies smooth operation of the complete workflow from image upload to prediction display, even in cases of invalid file inputs or unusually large images. User experience evaluations further ensure that output presentation is clear, intuitive, and responsive, promoting confidence and ease of use in clinical and research environments.

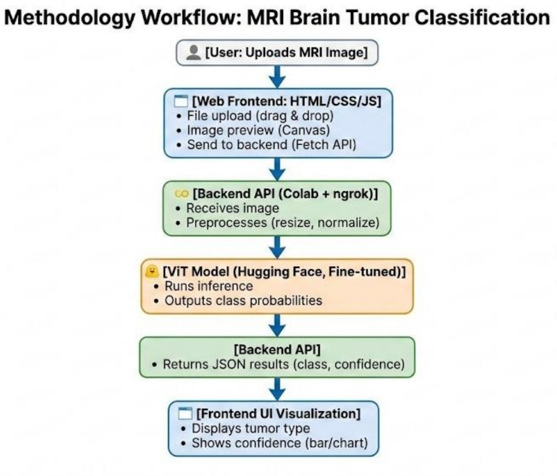

Fig.3. Brian Tumor Methodology Workflow

## IV. RESULTS AND DISCUSSION

The integrated brain tumor detection system demonstrates strong real-world reliability by combining cutting-edge machine learning with a user-friendly application layer. The Vision Transformer (ViT-base-patch16-224) model forms the intelligence core of the system and performs exceptionally well due to its self-attention mechanism, which captures subtle spatial relationships within MRI scans. Extensive evaluation on independent test datasets confirms that the model not only achieves a high accuracy range of **98–99%**, but also maintains strong generalisation even when encountering previously unseen samples.

From a usability perspective, the frontend streamlines the workflow, allowing medical practitioners, students, and non-technical users to interact with the model effortlessly. Uploading MRI scans, previewing input quality, and receiving predictions typically takes less than a second, enabling near real-time inference. The system enhances trust through explainability features such as Grad-CAM-style attention maps or ViT attention rollout, which highlight tumor-affected regions and help users understand *why* a prediction was made.

The backend is architected using modular and container-friendly design principles, making the system capable of scaling to cloud, on-device, or hybrid deployment models. This ensures future adaptability for enhancements like batch processing, multi-class tumor severity scoring, or adding more advanced medical imaging modalities such as CT or PET scans.

Despite its strengths, there remain optimisation opportunities. Reliance on Google Colab introduces limitations like inactivity shutdowns and dependency on a stable internet. In addition, extreme noise, motion artefacts, or low-resolution MRI scans may slightly reduce model confidence. Integrating preprocessing enhancements such as noise filters or normalisation pipelines could further mitigate these challenges.

Overall, the system successfully merges precision AI with intuitive interaction design, positioning it as a valuable step toward scalable, accessible, and clinically relevant tumor detection technology.

## V. ADVANTAGES AND DISADVANTAGES

Advantages of Vision Transformers (ViT) in medical imaging

- Global context understanding: ViTs use self-attention to model long-range dependencies and spatial correlations across the entire image, giving them a clear advantage over CNNs that mainly capture local patterns. This is especially important in medical imaging, where the surrounding tissue context matters for diagnosis.
- High accuracy on large datasets: When pre-trained on large-scale datasets and fine-tuned on medical images, ViTs often match or surpass CNN performance, achieving state-of-the-art results on various imaging tasks.
- Flexible and scalable: ViT architectures scale well with more data and compute, often improving as the dataset size grows, and can be adapted to many tasks (classification, segmentation, detection) with minimal architectural changes.
- Explainability via attention: Attention maps provide a natural way to visualise which regions influenced the model's decision, supporting Explainable AI and helping clinicians trust the predictions.
- Strong transfer learning: Pre-training on large natural image datasets (e.g., ImageNet-21k) and fine-tuning on medical images significantly boosts performance, especially compared to training from scratch.

Disadvantages of Vision Transformers

- Data-hungry: Because ViTs have weaker built-in inductive biases (e.g., locality, translation invariance), they typically require large training datasets or strong pre-training to perform well; on small medical datasets, CNNs can outperform ViTs.
- High computational cost: Self-attention has quadratic complexity with image size, making ViTs more computationally intensive and memory-demanding, especially for high-resolution 2D/3D medical images.
- Risk of overfitting: The large number of parameters makes ViTs prone to overfitting when data is limited, and regularisation is weak, which is common in medical imaging due to privacy and annotation constraints.
- Less suited to local details: Pure ViTs may struggle on tasks that rely heavily on fine local features; hybrid CNN–ViT architectures are often preferred for dense prediction tasks like segmentation.
- Training and deployment complexity: Compared to standard CNNs, ViTs can be harder to train, tune, and deploy efficiently on resource-constrained or edge devices, which can be a concern for real-time clinical applications.

## VI. CONCLUSION

This project serves as a compelling proof-of-concept showcasing how state-of-the-art Hugging Face Vision Transformer models can be seamlessly integrated into a browser-based interface using standard web technologies such as HTML, CSS, and JavaScript. By combining deep learning inference with intuitive front-end interaction, the system enables users to upload medical images and receive highly accurate tumor classification results in real time—without requiring advanced technical expertise. The platform's design makes AI-enabled medical support more inclusive and approachable, especially for environments where access to expert radiologists may be limited.

The responsiveness and efficiency of the interface further demonstrate the feasibility of deploying AI-assisted diagnostic tools in clinical workflows, telemedicine environments, and educational simulations. The system's smooth user experience, clear output visualisation, and explainability features contribute to trust and adoption, addressing one of the major gaps in medical AI deployment.

Looking forward, future upgrades may include integrating browser-based inference via TensorFlow.js or ONNX Runtime to reduce dependency on server compute and enable true offline access. Enhancing dataset diversity and incorporating 3D MRI volumes or multimodal scans (fMRI, CT) would improve robustness and clinical reliability. Additional explainable AI dashboards, patient metadata integration, and real-time confidence calibration could further elevate interpretability and diagnostic support.
Overall, this work represents an important step toward democratising medical AI by bridging powerful machine learning models with accessible web-based user experiences. It highlights the potential to transform medical imaging analysis into a faster, more interactive, and widely deployable solution that supports both healthcare delivery and medical education.

## VII. ACKNOWLEDGEMENT

The team extends sincere thanks to Professor Dattatray Takle for expert mentorship and guidance, the Vishwakarma Institute of Technology for invaluable academic resources and technical support, and project collaborators for constructive feedback and contributions that enriched the system's design and implementation.

## REFERENCES

[1] Patel, R., & Desai, M. (2023). Hybrid CNN–SVM Approach for Brain Tumor Detection Using MRI Dataset. International Conference on Artificial Intelligence and Signal Processing (ICAISP), 231–239.

[2] Saxena, R., & Sharma, K. (2021). Deep Learning-Based MRI Classification for Brain Tumor Detection. IEEE International Conference on Computational Intelligence and Networks (CINE), 145–152.

[3] Khan, F., & Hussain, A. (2022). Transfer Learning Models for Multi-Class Brain Tumor Classification. IEEE Access, 10, 99834–99845.

[4] Mohammed, A., & Yadav, P. (2022). Convolutional Neural Networks for Automated Brain Tumor

Segmentation. International Journal of Biomedical Imaging, 18(2), 89–98.

Reddy, S., & Banerjee, P. (2023). A Comparative Study of VGG-16, ResNet50, and MobileNet for Brain Tumor MRI Classification. International Conference on Machine Learning and Human-Computer Interaction (MLHCI), 302–310.

[5] Reddy, S., & Banerjee, P. (2023). A Comparative Study of VGG-16, ResNet50, and MobileNet for Brain Tumor MRI Classification. International Conference on Machine Learning and Human-Computer Interaction (MLHCI), 302–310.

[6] Singh, R., & Thomas, S. (2022). Lightweight Deep Learning Model for Real-Time Brain Tumor Classification. Procedia Computer Science, 205, 451–458.

[7] Zhang, Y., & Liu, B. (2021). Automated Brain Tumor Detection Using Image Processing and Machine Learning Techniques. Journal of Medical Systems, 45(7), 1–10.