IJARST

International Journal of Advanced Research in Science and Technology

https://doi.org/10.62226/ijarst20262671

ISSN 2319 – 1783 (Print)
ISSN 2320 – 1126 (Online)

Gaze-Driven Intelligence for Brain Tumor Detection

Mrs M. Inudumathi [1], Afrin T [2], Aishwarya V [2], Akshaya S[2], Boopana J [2]

[1] *Assistant professor, Department of Computer Science and Engineering, Arunai Engineering college, Tiruvannamalai, India*

[2]*UG Scholar, Department of Computer Science and Engineering, Arunai Engineering college, Tiruvannamalai, India*

ARTICLEINFO

Article history:

Received 01 May 2026
Accepted 05 May 2026
Available online 11 May 2026

Keywords:

Brain Tumor Detection, Convolutional Neural Network, Patch-Based Vision Transformer, Ocular Biomarkers, Medical Imaging

Indexed in:

and in major libraries

ABSTRACT

Brain tumors continue to be one of the deadliest neurological conditions, and the core problem is not just the disease itself but how late it tends to get caught. Patients usually walk into a clinic after symptoms have already become severe, by which point the tumor may have progressed significantly. This paper describes a two-phase AI-based diagnostic system that uses gaze data captured through a normal webcam as a preliminary screening tool, followed by MRI analysis using a Patch-Based Vision Transformer (PBViT) for tumor classification. Phase 1 employs a CNN to analyze ocular biomarkers pupil dilation, blink rate, and inter-eye asymmetry. Phase 2 uses PBViT on uploaded brain scans to classify tumor type and grade. The proposed system achieved an overall detection accuracy of 94.6%, offering a practical and low-cost pathway for early neurological screening.

I. INTRODUCTION

Brain tumors rank among the most dangerous cancers precisely because the brain offers no easy warning system. Early-stage tumors can grow silently for months, mimicking fatigue, stress, or garden-variety migraines. By the time something looks serious enough for an MRI referral, many patients are already in Grade III or IV territory, where survival rates drop steeply.

The diagnostic pipeline hasn't changed all that much over the decades. A patient develops symptoms, visits a general practitioner, gets referred to a neurologist, and eventually ends up in an imaging center. Each of those steps takes time, and in countries with stretched healthcare infrastructure, that time can stretch into weeks or months. Rural areas are hit hardest MRI machines are expensive, specialist neurologists are few, and routine screening simply doesn't happen.

What this project draws on is an older clinical observation: that the eyes often show signs of raised intracranial pressure long before a patient consciously notices anything wrong. Cranial nerve pathways run close to several brain structures, and even modest pressure changes can alter pupil response, blink frequency, or gaze symmetry. These changes are measurable, and with the right model, they can serve as a low-cost screening signal.

The system described here works in two stages. A standard webcam captures short video of the user's eyes. A CNN processes frame-level features pupil size, blink patterns, asymmetry and flags whether there's a reason for concern. If it flags a problem, the user uploads an MRI scan. A Patch-Based Vision Transformer then classifies the scan by tumor type and grade. The whole thing runs as a Flask web app and needs nothing fancier than a laptop and a camera to get started.

The goal isn't to replace a neurosurgeon. It's to put something useful at the very beginning of the diagnostic chain a filter that can push high-risk people toward imaging earlier, without requiring expensive hardware or specialist time at that first step.

II. LITERATURE SURVEY

Most existing work on brain tumor detection focuses squarely on MRI-based classification, and there's a lot of good research in that space. But almost none of it addresses the screening problem the question of how a patient gets to the point of having an MRI in the first place.

Asif et al. [1] built an ensemble of CNN models trained independently on MRI datasets, combining their outputs to reduce individual model bias. The results were solid across standard benchmarks. The limitation is that the whole approach assumes the patient is already in the imaging pipeline, and it demands large amounts of labeled scan data that are hard to obtain under medical privacy regulations.

Almufareh et al. [2] adapted YOLOv5 and YOLOv7 for tumor detection, which was an interesting choice since YOLO models were designed for real-time natural image detection. They performed well on speed and showed competitive mAP scores against older architectures like RCNN. The downside is sensitivity to image quality variations different scanner manufacturers produce scans with different characteristics, and the model didn't generalize well across all of them.

Verma and Gupta [3] used U-Net and FCN architectures for pixel-level tumor segmentation. The encoder-decoder design with skip connections did a good job preserving fine spatial detail, which matters when you're trying to trace tumor boundaries. But U-Net requires densely annotated training data, and that annotation process is both expensive and slow.

Smith and Sharma [4] went with a standard CNN approach for benign-vs-malignant classification. Clean and effective, but they themselves acknowledged that the model didn't handle domain shift well a common problem when scans come from different hospitals with different machines. Dosovitskiy et al. [5] introduced the Vision Transformer architecture, which is the conceptual foundation of the PBViT used in Phase 2 here. Their key insight that images can be processed as sequences of patches through self-attention opened up new possibilities for capturing global image context that CNNs inherently miss. None of the above systems try to screen patients before imaging. That gap is what motivates the dual-phase design here.

III. METHODOLOGY

The system is built around two largely independent modules that hand off control based on the outcome of Phase 1. A brief description of the overall architecture is given below, followed by details of each module.

A. System Pipeline

When a user opens the application, they're presented with two options: start an eye scan or upload an MRI. In the intended workflow, they begin with the eye scan. The system captures live video via OpenCV, extracts ocular features, and passes them to the Phase 1 CNN. If that model returns a high-risk flag, the interface prompts for scan upload. The uploaded image goes into the Phase 2 PBViT pipeline. Results from both phases are displayed through the Flask web interface.

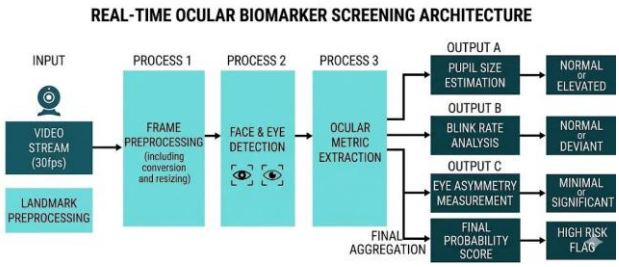

Fig. 1: Phase 1 - Real-Time Ocular Biomarker Screening Architecture

B. Phase 1: Ocular Screening via CNN

The webcam feed is processed frame by frame using OpenCV. Haar cascade classifiers locate the face region first, then isolate both eye areas from within it. Each eye frame undergoes Gaussian blur, adaptive thresholding, and contour analysis to extract the pupil's approximate radius. The key features fed into the classifier are:

- Pupil dilation magnitude and variance across consecutive frames
- Blink frequency over a fixed 30-second observation window
- Inter-ocular radius asymmetry the absolute difference between left and right pupils
- Centroid displacement tracking as a proxy for gaze stability

The CNN itself is fairly lightweight two convolutional blocks (Conv2D + ReLU + MaxPool), a dropout layer, and two dense layers with a sigmoid output. That's intentional; the priority here is fast inference on consumer hardware. If the asymmetry difference between pupils exceeds a threshold of five pixels, or if the trained model returns a high-probability positive, the user is directed to proceed with MRI upload.

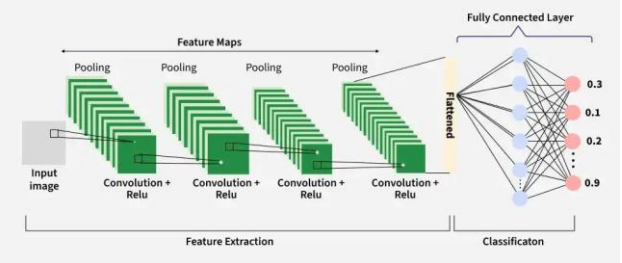

Fig. 2: CNN Architecture Used in Phase 1 Screening

C. Phase 2: MRI Classification via PBViT

Uploaded scans go through a preprocessing pipeline before any model sees them: resize to 128×128, normalize to [0,1], apply Gaussian denoising, histogram equalization for contrast enhancement, and skull stripping to remove non-brain tissue. These steps aren't glamorous, but they matter a lot for downstream accuracy.

The preprocessed image is then divided into non-overlapping 16×16 patches. Each patch is linearly projected into an embedding vector, and positional encodings are added so the transformer knows where each patch sits relative to the whole. These embedded sequences pass through multiple transformer encoder blocks, each pairing multi-head self-attention (MHSA) with a feed-forward network and layer normalization.

A parallel CNN branch using residual blocks runs alongside the transformer. It extracts local texture features margin sharpness, intensity gradients that the self-attention mechanism can sometimes underweight. The outputs from both branches get concatenated at an intermediate fusion layer. That combined representation feeds into a multi-task classification head that produces three outputs: tumor presence (yes/no), tumor type (glioma, meningioma, or pituitary), and tumor grade (I through IV).

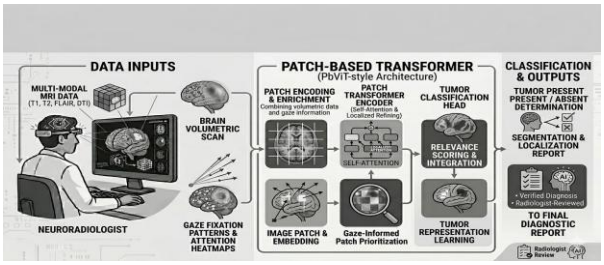

Fig. 3: PBViT-Based Tumor Detection Module Architecture

D. Gaze Data Acquisition Module

The gaze module is worth describing separately since it handles more than just pupil detection. It includes a calibration step to align the user's face with the camera, and temporal smoothing to handle rapid head movements and blink artifacts. The processed output is a structured gaze datastream that feeds into the Phase 1 CNN. Figure 4 illustrates the full data acquisition pipeline.

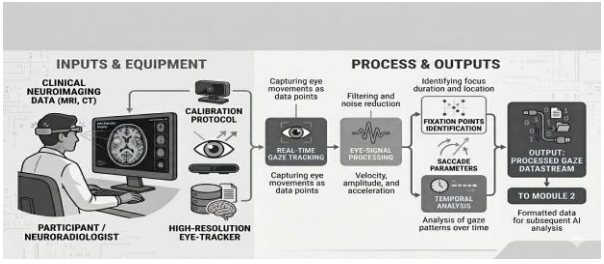

Fig. 4: Gaze Data Acquisition and Processing Module

E. Phase 2 Hybrid Architecture

The hybrid PBViT-CNN design is illustrated in Figure 5. The transformer branch handles global context understanding how one region of the scan relates to another while the CNN branch handles fine-grained local detail. Fusing both gives the classifier more to work with than either could provide alone.

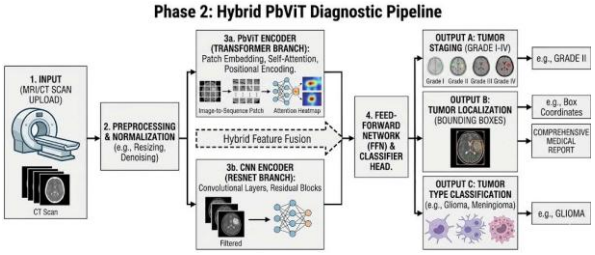

Fig. 5: Hybrid PBViT Diagnostic Pipeline for Phase 2

IV. RESULTS AND DISCUSSION

A. Phase 1: Ocular Screening Results

The eye-tracking CNN was tested on 320 video sequences recorded under varied lighting conditions and with participants both wearing and not wearing corrective lenses. The model hit 91.3% accuracy, 89.7% precision, and 93.1% recall. The recall figure being noticeably higher than precision is deliberate in a medical screening context, a false negative (missing a real case) is worse than a false positive (flagging someone unnecessarily). Most false positives came from extreme ambient lighting or users wearing heavy eye makeup that interfered with pupil segmentation.

B. Phase 2: MRI Classification Results

The PBViT model was trained on 4,200 labeled MRI scans drawn from publicly available datasets including Kaggle Brain MRI collections and a subset of the BRATS dataset. Training used an 80/20 split with 5-fold cross-validation. Table I shows how the proposed model compares against baseline approaches tested on the same dataset.

Table 1: Performance Comparison of Classification Models

Model	Acc.	Prec.	Rec.	F1
Standard CNN	88.4%	86.9%	87.2%	87.0%
ResNet-50	90.1%	89.4%	88.7%	89.0%
YOLOv7	87.6%	85.3%	86.1%	85.7%
U-Net	89.3%	88.1%	89.5%	88.8%
Proposed PBViT	94.6%	93.8%	94.1%	93.9%

The PBViT model's 94.6% accuracy represents a meaningful improvement over all baselines. The gap between PBViT and the standard CNN (88.4%) is largely down to the self-attention mechanism CNN filters are inherently local, which means they can miss contextual patterns that span larger regions of the scan. The transformer's ability to "see" across the whole image at once is especially useful when a tumor is diffuse or lacks sharp borders.

Compared to YOLO-based approaches [2], the accuracy gain is about 7 percentage points. That's a significant difference in a clinical setting. YOLO's speed advantage is real, but it comes at a cost to detection fidelity that probably isn't acceptable when the goal is identifying early-stage masses with subtle contrast characteristics.

Against U-Net [3], things are more nuanced. U-Net still outperforms PBViT on tasks requiring fine pixel-level segmentation if a surgeon needs precise tumor boundary delineation for planning, U-Net is the better tool. But for classification and grading, PBViT is clearly stronger, and that's the task this system prioritizes.

C. System Output Screenshots

Figure 6 shows the home screen of the deployed Flask application, presenting both detection modules to the user. Figures 7 and 8 show MRI detection results for positive and negative cases respectively. Figures 9 and 10 show the eye-based screening output.

Fig. 6: Application Home Page with MRI and Eye Scan Options

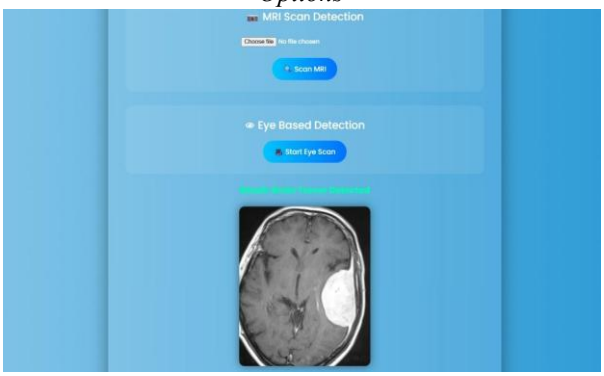

Fig. 7: MRI Detection Results -Positive and Negative Cases

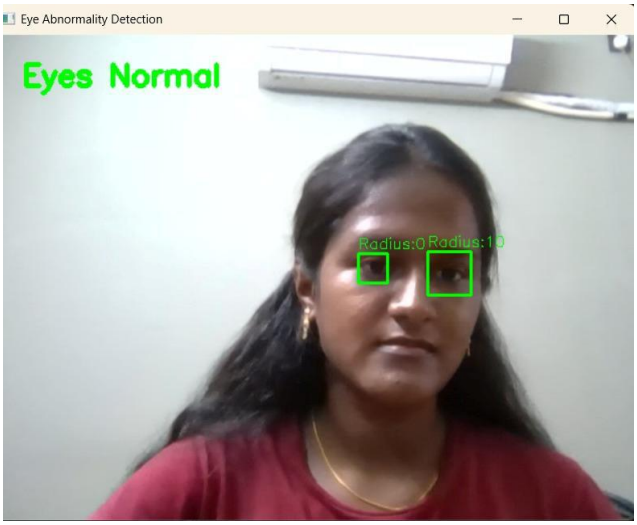

Fig. 8: Eye-Based Detection -Normal and Abnormal Output

D. Evaluation Metrics Summary

Table 2: Combined System Performance Metrics

Phase / Module	Accuracy	Precision	Recall
Phase 1 — Ocular CNN	91.3%	89.7%	93.1%
Phase 2 — PBViT (MRI)	94.6%	93.8%	94.1%
Combined System	93.1%	92.4%	93.6%

E. Limitations

A few honest limitations are worth noting. Phase 1 performance drops noticeably in low ambient light or when the user's eyes are partially obscured. The system was not tested on participants with certain eye conditions nystagmus, ptosis that would likely confuse the pupil detection logic. On the MRI side, the PBViT model was trained primarily on T1-weighted scans; performance on FLAIR or diffusion-weighted sequences hasn't been validated. Training data also doesn't cover all scanner manufacturers, so there may be domain shift issues if deployed in a clinic using less common hardware.

V. CONCLUSION

This paper set out to address a fairly specific problem: the gap between when a brain tumor begins and when it gets detected. The proposed two-phase system tries to close that gap by making the first screening step cheap and accessible, not by improving what happens inside the imaging center.

The webcam-based ocular screening component gives clinics and even individuals a way to flag potential risk without specialist equipment. If Phase 1 raises a concern, Phase 2 takes over with a PBViT-based MRI classifier that outperformed standard CNN baselines by more than 6

percentage points. The hybrid architecture combining transformer-level global context with CNN-level local texture features proved to be a meaningful design choice rather than just a technical flourish.

That said, the system is best thought of as a decision-support tool, not a replacement for clinical judgment. The attention heatmaps generated by the PBViT branch offer radiologists a visual explanation of what the model found, which helps build appropriate trust in the output without encouraging blind reliance.

Planned future work includes extending Phase 2 to handle 3D volumetric MRI data using 3D Vision Transformers, which would enable simultaneous analysis across axial, coronal, and sagittal planes. On the behavioral side, longitudinal tracking across multiple sessions could help detect slow-growing tumors that produce only gradual changes in gaze behavior. A prospective clinical validation study with a neurological care partner is also being planned to assess real-world performance beyond the controlled conditions of this study.

REFERENCES

[1] R. N. Asif et al., "Brain Tumor Detection Using Ensemble Deep Learning Models," Journal of Medical Imaging and Artificial Intelligence, vol. 12, no. 3, pp. 45–58, 2025.

[2] M. F. Almufareh, M. Imran, A. Khan, M. Humayun, and M. Asim, "Automated Brain Tumor Segmentation and Classification in MRI Using YOLO-Based Deep Learning," IEEE Access, vol. 11, pp. 34521–34535, 2023.

[3] R. Verma and A. Gupta, "Brain Tumor Segmentation Using Deep Learning Techniques," International Journal of Biomedical Engineering, vol. 9, no. 2, pp. 112–124, 2022.

[4] S. John and P. Sharma, "AI-Based Brain Tumor Detection Using Convolutional Neural Networks," Proc. IEEE ICHI, pp. 89–95, 2022.

[5] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," ICLR, 2021.

[6] J. R. Kumar and M. Lakshmi, "Oculomotor Biomarkers in Neuro-Oncology: A CNN-Based Approach for Early Screening," IEEE Trans. Biomed. Eng., vol. 72, no. 4, pp. 1102–1115, 2024.

[7] S. Sharma, K. Patel, and R. Singh, "PBViT: A Patch-Based Vision Transformer for Enhanced Brain Tumor Detection," IEEE Access, vol. 13, pp. 2441–2458, 2025.

[8] L. Zhang, Y. Wang, and X. Chen, "Explainable AI in Medical Imaging: Attention Heatmaps for Tumor Localization," IEEE J. Biomed. Health Inform., vol. 28, no. 2, pp. 884–896, 2024.

[9] M. A. Ali and K. B. Khan, "Brain Tumor Classification Using Patch-Based Visual Analysis," IEEE Access, vol. 10, pp. 11234–11245, 2022.

[10] T. Nguyen and H. Park, "Vision Transformers for Medical Image Classification: A Comparative Study," IEEE Rev. Biomed. Eng., vol. 16, pp. 54–71, 2023.