

RAG-Powered AI System With Explainability For The Detection And Classification Of Brain Tumor

Umme Amara Shariff¹, Gowtham N D², Syeda Zahara Sultana³, Vedhashree MY⁴, SoumyaB J⁵

^{1, 2, 3, 4} Dept of ISE

⁵ Assistant Professor, Dept of ISE

^{1, 2, 3, 4, 5} SAM MRIT, Many, India

Abstract- Brain tumor classification using MRI is important for medical image assessment. This paper presents a RAG-powered AI system that integrates deep-learning classification, explainable AI, and medical knowledge retrieval. The system aims to classify brain MRI scans and provide understandable explanations supported by relevant medical information. The proposed framework combines image preprocessing, classification, visual explanation, and retrieval-augmented generation through a user-friendly interface.

Keywords- Cloud-based System, GPS, Mobile Technology, Traffic Violation.

generates text without retrieval and source checks. The project addresses this integration problem by proposing a pipeline that connects classification, visual attribution, medical-information retrieval, and explanation generation.

The objectives are to: (1) provide an interface for uploading brain MRI images; (2) classify images into the project categories; (3) explore XAI methods to visualize model attention or feature contribution; (4) retrieve relevant information from a curated knowledge base; (5) generate concise explanations that can be traced to retrieved sources; and (6) evaluate the classifier, retrieval quality, explanations, and interface after implementation.

I. INTRODUCTION

Brain tumors comprise different types of abnormal growths within or around the brain. MRI is commonly used in clinical assessment because it provides detailed images of soft tissues without using ionizing radiation. Interpreting these images requires specialist expertise, and image appearance can vary with tumor type, size, location, acquisition protocol, and image quality. Computer-aided analysis is therefore an active research area, although automated predictions must be validated carefully before they can be relied upon in practice.

Convolutional neural networks and transfer-learning models can learn visual patterns from labelled images and assign an input scan to a defined class. However, a class label alone may not show which image regions influenced the model. Explainable AI methods can provide visual or feature-based evidence, while retrieval-augmented generation can retrieve relevant passages from a trusted document collection before a language model writes an explanation. This project proposes bringing these capabilities together in one workflow.

II. PROBLEM STATEMENT AND OBJECTIVES

A standalone classifier may return a category without an accessible explanation. A heatmap may indicate influential regions but does not, by itself, explain the medical meaning of those regions. Similarly, a language model may produce statements that are not supported by reliable evidence if it

III. LITERATURE SURVEY

Research on automated brain MRI analysis includes conventional machine-learning approaches, convolutional neural networks (CNNs), transfer learning, and ensembles. CNNs learn spatial patterns from image data, while transfer learning adapts features learned from larger image collections to a target dataset. Ensemble methods combine predictions from multiple models and may improve robustness, but their benefits depend on dataset quality, model diversity, and evaluation design.

Explainability research explores methods that help users inspect model behavior. Grad-CAM creates a class-specific heatmap from feature maps and gradients in a convolutional network. LIME approximates a model locally by perturbing input features, while SHAP uses contribution values to describe how features affect an output. These methods have different assumptions and limitations; a visually appealing heatmap should not be treated as proof that a prediction is medically correct.

Retrieval-Augmented Generation is a method in which a system searches an external collection of documents and supplies relevant passages to a language model before generation. In healthcare-oriented applications, retrieval can make explanations more traceable by linking generated text to source documents. The approach still depends on the quality and currency of the knowledge base, retrieval relevance, and

whether the generated explanation accurately reflects the retrieved evidence.

Research area	Contribution	Relevance to this project
CNN / transfer learning	Learns image features for multiclass classification.	Provides the proposed MRI classifier.
Ensemble methods	Combines predictions from multiple models.	Possible future baseline comparison.
Grad-CAM	Highlights image regions associated with a class score.	Adds a visual explanation layer.
LIME / SHAP	Estimates local or feature contributions.	Potential complementary XAI techniques.
RAG in healthcare	Retrieves documents to support generated text.	Grounds explanations in medical sources.

The project's literature review identifies integration as the main design direction: a classifier supplies a predicted category, XAI offers an image-level explanation, and RAG supplies contextual evidence for readable reporting. This is a project-defined research direction rather than a claim that no comparable system exists. The novelty and scope must be confirmed by examining the full text of relevant publications.

IV. PROPOSED METHODOLOGY

The workflow is organized into six stages: MRI input, preprocessing, classification, explainability, knowledge retrieval, and report presentation. Each stage has a defined role, and the final explanation should distinguish the classifier's output from statements found in retrieved medical sources.

4.1 IMAGE INPUT AND PREPROCESSING

The user selects a brain MRI image through the MedVision interface. Before inference, the image should be checked for supported file type and successful loading. Preprocessing may include resizing to the model input dimensions, conversion to the required channel format, and

normalization. Augmentation may be applied to training images when suitable, but evaluation images should be processed consistently without transformations that could invalidate comparisons. The exact image size, normalization values, and dataset split must be documented after implementation.

4.2 CLASSIFICATION MODULE

A CNN or transfer-learning architecture can be trained to map an MRI image to one of the project's four labels: No Tumor, Glioma, Meningioma, or Pituitary. The model learns patterns from labelled training images, and a validation set can be used to select hyperparameters and monitor overfitting. At inference time, the system should display the predicted class and, if the model provides calibrated or appropriately described scores, the associated confidence information. A score should not be described as a probability of disease unless calibration and interpretation have been validated.

4.3 EXPLAINABLE AI MODULE

Grad-CAM is a candidate technique for producing a heatmap that emphasizes regions contributing to a selected class score. The heatmap can be displayed alongside the original MRI to support inspection of the model's behavior. LIME or SHAP may be evaluated as complementary methods if they are compatible with the selected architecture and input representation. Explanations must be interpreted cautiously: highlighted regions can reflect artifacts or dataset bias, and XAI output does not establish a clinical diagnosis.

4.4 KNOWLEDGE BASE AND RETRIEVAL

The knowledge base should contain identifiable, reputable medical documents relevant to brain tumors and MRI. Each document should be cleaned and divided into passages of manageable length. Metadata such as document title, organization or publisher, date, and source location should be retained so that the final answer can cite its supporting material. Only documents actually included in the implementation should be described as part of the knowledge base.

In a typical RAG pipeline, passages are converted into vector embeddings and indexed using a vector search tool such as FAISS or ChromaDB. A user request or classifier result is used to retrieve semantically relevant passages. The generation component receives the predicted class, a summary of the visual explanation, and the retrieved passages. Its instructions should require it to use the supplied evidence,

avoid unsupported claims, and indicate when the retrieved material is insufficient.

4.5 EXPLANATION AND REPORT GENERATION

The report interface can combine the original image, predicted category, XAI visualization, a concise explanation, and source details. The output should separate three kinds of information: the model's prediction, the visual attribution generated by the XAI method, and the medical context stated by retrieved sources. The interface should include a clear notice that the prototype is intended for research and educational decision support, not independent diagnosis or treatment recommendation.

4.6 EVALUATION PLAN

Classification should be assessed on a held-out test set using accuracy, per-class precision, recall, F1-score, and a confusion matrix. Sensitivity and specificity may also be reported when the class definition and evaluation design make them appropriate. Retrieval should be checked for relevance and source traceability; generated explanations should be reviewed for factual support, omissions, and unsupported statements. Interface evaluation may examine ease of use and response time. All values must be calculated from actual tests, not estimated or invented.

V. SYSTEM ARCHITECTURE

The system consists of an input interface, image preprocessing module, deep-learning classifier, explainability module, medical knowledge base, retrieval engine, and reporting interface. The MRI image is processed by the classifier, while explainability and retrieval provide complementary information for the final report.

The expected data flow is: (1) upload an MRI image; (2) validate and preprocess the image; (3) run the classifier; (4) generate a visual explanation; (5) retrieve relevant passages from the curated knowledge base; (6) generate an explanation with source references; and (7) present the image, prediction, explanation, and sources in the interface. If retrieval returns weak or unrelated evidence, the system should avoid confident medical statements and indicate that supporting information was not found.

VI. IMPLEMENTATION

The MedVision interface provides a dashboard with Brain MRI and Chest X-ray analysis options. The Brain MRI page includes an image-upload interface and an option to

initiate analysis. The workflow screen displays four brain MRI categories: No Tumor, Glioma, Meningioma, and Pituitary, alongside the image-to-prediction.

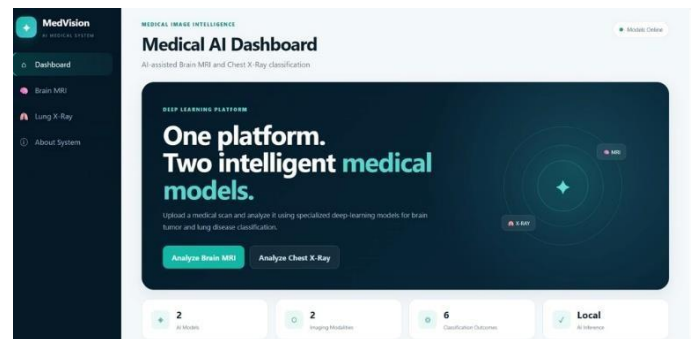


Fig 1. Medical AI Dashboard

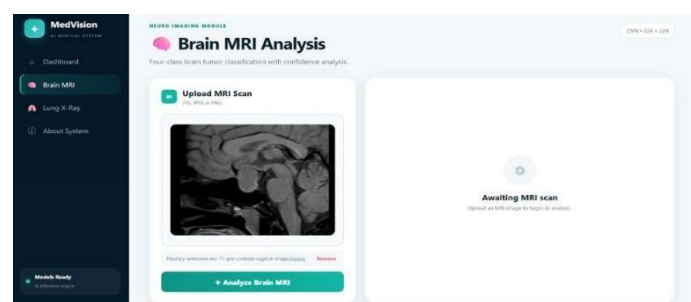


Fig 2. Brain MRI Analysis Interface

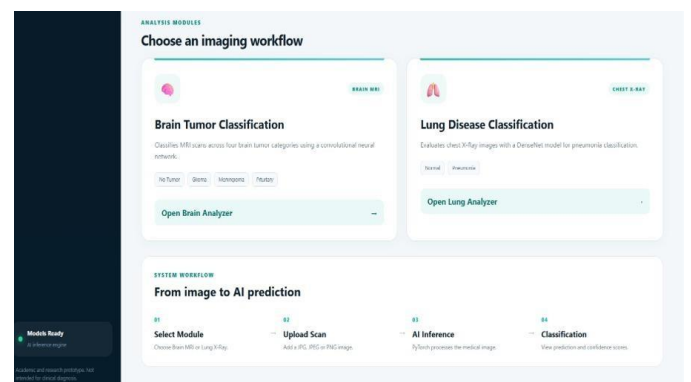


Fig 3. Imaging workflow and classification modules

VII. RESULTS AND DISCUSSION

The available screenshots demonstrate the dashboard, the brain MRI upload interface, and the intended classification workflow. They document the user-facing structure and the four displayed categories. They do not show a completed MRI prediction, measured classification performance, retrieved source passages, or a generated explanation. Accordingly, this paper reports interface-level observations only and does not claim a measured accuracy, precision, recall, F1-score, or clinical benefit.

After the full pipeline is implemented, the results section should include the test-set size and class distribution, a confusion matrix, per-class metrics, representative correctly and incorrectly classified images, example Grad-CAM visualizations, retrieved passages, and sample generated explanations with source references. The discussion should examine errors and limitations, including class imbalance, image quality, data leakage, model bias, irrelevant retrieval, and unsupported generated text. Any numerical result must be reproduced from recorded experiments.

VII. LIMITATIONS AND FUTURE WORK

The current evidence supplied for this paper is limited to interface screenshots and the proposed workflow. The exact dataset, trained architecture, preprocessing configuration, vector index, language model, and experimental measurements must be confirmed from the working implementation before being stated as completed. Future work includes completing the classifier and RAG pipeline, curating and documenting medical sources, testing XAI outputs, measuring retrieval relevance and answer groundedness, and evaluating the interface with users. External validation and qualified clinical review would be required before considering any real-world clinical use.

VIII. CONCLUSION

This paper presents the design of a RAG-powered brain MRI classification system that aims to combine deep learning, explainable AI, and retrieval-grounded medical information. The proposed workflow is intended to make outputs more interpretable by pairing a predicted class with visual attribution and source-supported text. The supplied MedVision screenshots demonstrate the dashboard, MRI upload page, and intended four-class workflow. Full implementation and experimental validation remain necessary before making claims about predictive performance, explanation quality, or clinical usefulness. The system should be treated as a research prototype and not as a substitute for professional medical assessment.

IX. ACKNOWLEDGMENT

The authors thank the management of Mysuru Royal Institute of Technology, the Department of Information Science and Engineering, and their project guide, Prof. Sowmya B. J., for their guidance and academic support.

REFERENCES

- [1] Kakon et al., “Explainable Deep Ensemble Meta-Learning Framework for Brain Tumor Classification,” *Cancers*, 2025. Verify full publication details.
- [2] Tandel et al., “Multi-Class Brain Tumor Grades Classification Using Majority Voting and XAI,” 2025. Verify full publication details.
- [3] M. et al., “Enhancing Brain Tumor Detection Using Grad-CAM with ResNet50,” *BMC Medical Imaging*, 2024. Verify full publication details.
- [4] Weinert and Rauschecker, “Retrieval-Augmented Generation: A Radiology-Specific Approach,” 2025. Verify full publication details.
- [5] Amugongo et al., “Retrieval Augmented Generation in Healthcare: A Systematic Review,” 2025. Verify full publication details.