Automated Brain Tumor Diagnosis from Magnetic Resonance Imaging using Convolutional Neural Networks: A MATLAB-Based Approach

Authors: Vignesh Chivirala, Aprameya Goud, Sreeja Muthurla, Vijay Teja Asagunda

Abstract

Brain tumors represent a significant cause of morbidity and mortality, making early and accurate diagnosis essential for effective treatment planning and patient outcomes. Manual interpretation of Magnetic Resonance Imaging (MRI) scans is a complex, time-consuming process susceptible to human error and inter-observer variability.[1] This report details a comprehensive framework for the automated classification of brain tumors using Convolutional Neural Networks (CNNs). The objective is to develop a robust, high-performance diagnostic system implemented within the MATLAB environment. This paper outlines the complete methodology, beginning with the curation of essential public datasets (including the Figshare 3-class and Kaggle 4-class repositories) [3] and detailing a rigorous, mandatory preprocessing pipeline, with special emphasis on skull stripping.[5] A deep mathematical analysis of the core CNN components—including convolutional layers, ReLU activation, and max-pooling—is presented. The report proposes a high-performance model based on ResNet transfer learning, which benchmark studies indicate can achieve classification accuracy and F1-scores approaching 99%.[6] The system's viability as a clinical support tool is evaluated, concluding that CNNs can substantially augment radiological workflows. Finally, the report addresses the primary limitations and future challenges, such as the "black box" problem of model interpretability, data generalization, and patient privacy.[8]

1. Introduction

1.1. The Clinical Challenge of Brain Tumors

Brain tumors, characterized by the uncontrolled and abnormal proliferation of cells within the brain, represent a significant global health burden.[10] They are a major cause of cancer-related mortality, particularly in malignant forms such as glioblastoma.[7] The prognosis for malignant brain tumors is often poor [13], with the five-year survival rate for cancerous central nervous system (CNS) tumors estimated at only 34% for men and 36% for women.[14]

This stark prognosis underscores the critical importance of rapid, accurate, and early diagnosis. The ability to detect a tumor at its nascent stage, correctly identify its type, and precisely delineate its boundaries is paramount for determining the optimal treatment strategy, which may include surgery, radiotherapy, or chemotherapy.[13] Timely and effective treatment planning, enabled by an accurate diagnosis, is the most crucial factor in improving patient outcomes and survival rates.[10]

1.2. The Role of Magnetic Resonance Imaging (MRI)

In the field of neuro-oncology, neuroimaging is the cornerstone of diagnosis, and Magnetic Resonance Imaging (MRI) is the undisputed gold-standard modality.[13] Unlike Computed Tomography (CT) or X-rays, MRI is a non-invasive technique that does not use ionizing radiation.[10] Its primary advantage lies in its unparalleled soft-tissue contrast and high spatial resolution, which provide detailed anatomical visualization of brain structures.[10]

MRI is indispensable in all phases of patient management. It is used for initial diagnosis and characterization, pre-surgical planning (often with functional MRI (fMRI) to map critical brain areas), and post-treatment follow-up to monitor for treatment response or tumor recurrence.[13] Advanced MRI sequences, such as T1-weighted contrast-enhanced scans, T2-weighted scans, and Magnetic Resonance Spectroscopy (MRS), can reveal not only the morphology of a tumor but also its metabolic, cellular, and microstructural properties.[13]

1.3. Limitations of Manual Radiological Diagnosis

While MRI provides rich data, the interpretation of this data is a significant bottleneck. A single patient's scan can consist of thousands of individual images that must be meticulously examined by a neuroradiologist. This manual analysis, despite being the current standard, is a complex, time-consuming, and laborious task [10] and is subject to several well-documented human-factor limitations:

- **Subjectivity and Variability:** The appearance of tumors can be highly heterogeneous, varying in size, shape, texture, and location. This complexity leads to inter-observer variability, where different experts may arrive at different conclusions based on a subjective interpretation of the same image.[1]
- **Radiologist Fatigue:** Fatigue is a critical factor that directly and negatively impacts diagnostic accuracy.[1] Studies have shown that diagnostic error rates, which have been estimated to be as high as 30%, have not significantly improved in over 50 years, in part due to this human factor.[2] The cognitive load of interpreting complex scans, especially during long shifts or at night, demonstrably increases the rate of diagnostic errors.[21]
- **Cognitive Bias:** Radiologists, like all human experts, are susceptible to cognitive biases. For example, attribution bias can occur when a pre-existing clinical note (e.g., from a referring physician) primes the radiologist to "see" an expected, but incorrect, diagnosis, leading to a misinterpretation of the visual evidence.[21]

These limitations can lead to diagnostic delays or inaccuracies, which in turn can adversely affect patient outcomes.[1]

1.4. Deep Learning as a Solution in Neuro-Oncology

To address the challenges of manual diagnosis, Computer-Aided Diagnosis (CAD) systems have emerged as a powerful solution.[9] In particular, the field has been revolutionized by deep learning, a subfield of machine learning that utilizes multi-layered artificial neural networks.[24]

Convolutional Neural Networks (CNNs) are a specific class of deep learning models explicitly designed to process grid-like data, such as images.[25] CNNs have demonstrated state-of-the-

art performance in numerous medical imaging tasks.[8] Their primary strength is their ability to automatically learn relevant features from raw pixel data, bypassing the need for manual and subjective feature engineering.[23] A CNN can learn a hierarchy of patterns, from simple edges and textures in its initial layers to complex, abstract representations of tumor morphology in its deeper layers.[1]

By providing a rapid, objective, and accurate classification, a well-trained CNN acts as an indefatigable "second opinion".[1] These systems can enhance diagnostic accuracy [1], reduce human error [30], and streamline clinical workflows by prioritizing urgent cases [25], ultimately augmenting the capabilities of the human radiologist.

1.5. Project Objective

The primary objective of this projec is to design, implement, and evaluate a CNN-based system for the automated classification of brain tumors from MRI scans. The system will be developed using the MATLAB programming environment and its Deep Learning Toolbox. The goal is to create a robust classifier capable of distinguishing between the tumor—and a healthy "no tumor" class, providing a clinically relevant framework for both detection and typing.

---

2. Methodology

2.2. Dataset Acquisition and Selection

The foundation of any successful deep learning model is a large, high-quality, and well-annotated dataset.[31] For this project, a thorough review of public repositories was conducted to identify datasets suitable for a MATLAB-based classification task.

- **Dataset 1: Multi-Class Tumor Typing (Figshare)**
  - **Source:** The "brain tumor dataset" published by Jun Cheng, et al. on the Figshare repository.[3]
  - **Contents:** This dataset contains 3,064 T1-weighted contrast-enhanced MRI images sourced from 233 patients.[3]

- **Project Relevance:** This dataset is of exceptional relevance to this specific project. The data is provided in .mat files [33], and the accompanying readme file *explicitly includes MATLAB code* to convert these .mat files into standard JPG images.[3] This direct compatibility significantly streamlines the project's data-loading phase and validates the choice of MATLAB as the development environment.[37]

2.3. Image Preprocessing Pipeline

Raw MRI data is heterogeneous and unsuitable for direct input into a CNN. A rigorous, multi-step preprocessing pipeline is mandatory to standardize the data, remove confounding information, and optimize the model for learning pathological features.[12]

- **2.3.1. Skull Stripping (Brain Extraction)**
  - This is arguably the most critical and non-negotiable step in the pipeline.[5] Skull stripping is the process of algorithmically removing all non-brain tissues from the MRI scan, including the skull, scalp, dura mater, eyes, and fat.[5]
  - **Rationale:** These non-brain tissues have high signal intensities in an MRI and create strong, high-contrast edges. If they are not removed, the CNN will *not* learn to find tumors; it will learn to find skulls. The model will achieve high accuracy on the test set (which also contains skulls) but will have learned a useless, non-pathological feature. Any error at this stage—either including non-brain tissue or accidentally removing brain tissue—*cannot be rectified* by subsequent steps and will propagate, invalidating the entire model.[5]
  - **Method:** This process will be automated using established tools such as FMRIB's Brain Extraction Tool (BET) [47] or similar algorithms available in neuroimaging packages.
- **2.3.2. Image Normalization and Resizing**
  - **Normalization:** MRI signal intensities are not absolute; they can vary significantly between different scanners, patients, and acquisition protocols. To create a standardized input, Z-score normalization will be applied. For each image, the mean intensity of the *brain region* (post-skull stripping) will be subtracted from every pixel, and the result will be divided by the standard deviation.[12]
  - **Resizing:** CNN architectures require a fixed, uniform input size. All preprocessed, skull-stripped images will be resized to a standard dimension (e.g., 256×256 pixels) to match the input layer of the model.[31]
- **2.3.3. Data Augmentation**
  - Deep learning models are data-hungry, and performance is often limited by the size

of the training set.[31] Medical image datasets are often small, which increases the risk of overfitting (where the model "memorizes" the training data but fails to generalize to new images).[49]

- Data augmentation is a technique to artificially inflate the training set by creating a large number of new, plausible training samples.[22] These transformations will be applied randomly and on-the-fly *only* to the training data.
- **Techniques:** The augmentation pipeline will include a combination of affine image transformations [52]:
  - **Rotation:** Randomly rotating the image (e.g., $\pm$20 degrees).[47]
  - **Flipping:** Randomly flipping the image horizontally.[50]
  - **Scaling/Zooming:** Randomly zooming in on a portion of the image (e.g., 0-20% zoom).[52]
  - **Shifting:** Randomly shifting the image horizontally or vertically.[53]
- This process makes the model more robust to variations in position and orientation, significantly improving its ability to generalize to unseen clinical data.

2.4. Model Architecture Development in MATLAB

	Function / Module	Input	Output
1	Image Upload (`uigetfile`)	MRI Image	Raw Image
2	Preprocessing (`imnoise`, denoise)	Raw Image	Noise-free MRI
3	Skull Stripping	Preprocessed Image	Brain-only Image
4	Segmentation (`FCM`, `BIBO`)	Skull-stripped Image	Tumor Region
5	Feature Analysis	Segmented Region	Precision, Recall, Accuracy
6	GUI Display	Outputs from 2–5	User Visualization

- **2.4.2. Advanced Model: Transfer Learning**
  - This is the primary proposed model. Training a deep CNN from scratch requires an enormous amount of data. Transfer learning (TL) is a technique that circumvents this problem.[22]
  - **Concept:** TL leverages a model (e.g., VGG16, ResNet, GoogleNet) that has *already* been trained on a massive, general-purpose dataset like ImageNet (which contains millions of images).[18] The initial layers of this pre-trained model have already learned to be powerful, universal feature detectors for edges, textures, shapes, and patterns. These learned features are highly effective for medical imaging as well.[59]
  - **MATLAB Implementation:** The MATLAB Deep Learning Toolbox provides these pre-trained models.[37] The process is:
    1. **Load Model:** Load a pre-trained network, such as resnet18 or googlenet.
    2. **Freeze Layers:** "Freeze" the weights of the initial convolutional layers to prevent them from being altered during training. This retains their learned feature-extraction capabilities.
    3. **Replace Final Layers:** The final classification are removed. They are replaced with new, "unfrozen" layers customized for this task: a fullyConnectedLayer (with 4 outputs for our 4 classes), a softmaxLayer, and a classificationLayer.
    4. **Fine-Tuning:** The new, hybrid network is then trained ("fine-tuned") only on the brain tumor dataset. This approach requires significantly less data, trains much faster, and typically achieves a much higher classification accuracy than a model trained from scratch.

---

3. Mathematical Analysis

This section provides a foundational mathematical and conceptual explanation of the "CNN concept," deconstructing the components outlined in the methodology.

3.1. Classical Pipeline: Explicitly defines each step. The "intelligence" is in the hand-crafted math of the FCM or Rough Set algorithm.

CNN Approach : An end-to-end system. After preprocessing (like skull stripping), a single CNN would be trained to learn the features and classification/segmentation task directly from the pixels, rather than being guided by explicit clustering math.

---

4. Experiments and Results

This section outlines the experimental setup for validating the proposed model and presents the *expected* performance targets. These targets are synthesized from a comprehensive review of benchmark results in recent literature [6] and serve as the quantitative goal for this project.

4.1. Experimental Setup

- **Data Split:** The Kaggle 4-class dataset [4] will be partitioned into three distinct sets:
  - **Training Set (80%):** Used to train the model.
  - **Validation Set (10%):** Used during training to monitor performance on unseen data. This set is used to tune hyperparameters (like learning rate) and prevent overfitting by triggering "early stopping" if the validation loss stops improving.[7]
  - **Test Set (10%):** Held back and used *only once* at the very end of the project to provide a final, unbiased assessment of the model's generalization performance.[55]
- **Environment:** All experiments will be conducted in MATLAB R2021a (or newer) using the Deep Learning Toolbox [37] and accelerated on an NVIDIA T4 GPU or equivalent CUDA-enabled device.[1]

4.2. Evaluation Metrics

To move beyond simple accuracy, a comprehensive suite of classification metrics will be used to evaluate the model.[11] These metrics are all derived from the Confusion Matrix.

- **4.2.1. Confusion Matrix**
  - The confusion matrix is a table that provides a complete visualization of model performance by cross-referencing the *Actual Labels* with the *Predicted Labels*.[92] For the 4-class problem, this will be a 4×4 matrix.
  - The fundamental components are best understood in the binary context (e.g., Tumor vs. No Tumor) [93]:
    - **True Positive (TP):** Actual = Tumor, Predicted = Tumor. (A correct detection).
    - **True Negative (TN):** Actual = No Tumor, Predicted = No Tumor. (A correct

rejection).
    - **False Positive (FP) (Type I Error):** Actual = No Tumor, Predicted = Tumor. (A false alarm; leads to unnecessary follow-ups).
    - **False Negative (FN) (Type II Error):** Actual = Tumor, Predicted = No Tumor. (A missed detection; the most severe clinical error).
- **4.2.2. Metric Formulas**
    - Accuracy: The percentage of all correct predictions.

      $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$
      .96
      Note: Accuracy can be highly misleading in imbalanced datasets (e.g., if 99% of scans are "No Tumor," a model that predicts "No Tumor" every time has 99% accuracy but is clinically useless).93
    - Precision (Positive Predictive Value): The percentage of predicted positives that were correct. "Of all the times the model cried 'Tumor', how often was it right?"

      $$Precision = \frac{TP}{TP + FP}$$
      .96
      High precision is important for avoiding False Positives and unnecessary procedures.
    - Recall (Sensitivity): The percentage of actual positives that were correctly identified. "Of all the real tumors, how many did the model find?"

      $$Recall = \frac{TP}{TP + FN}$$
      .96
      High recall is the most critical metric for a screening tool, as it focuses on minimizing False Negatives (missed tumors).
    - F1-Score: The harmonic mean of Precision and Recall. It provides a single, balanced score.

      $$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN}$$
      .96
      The F1-score is the preferred metric for imbalanced medical datasets because it penalizes models that excel at only one of Precision or Recall.96

4.3. Expected Performance Analysis

The following table synthesizes benchmark results from multiple recent studies [6] to establish

performance targets. This clearly illustrates the expected gap between a baseline model (Custom CNN) and the advanced transfer learning model.

4.5. Visualization of Results

1. **Training & Validation Curves:** Plots of accuracy and loss for both the training and validation sets over each epoch.[7] A large, persistent gap between the training and validation curves would indicate overfitting.
2. **ROC Curve (Receiver Operating Characteristic):** For the binary (Tumor vs. No Tumor) task, the ROC curve will be plotted. This graph plots the True Positive Rate (Recall) against the False Positive Rate.[100] The **Area Under the Curve (ROC-AUC)** will be calculated; a value close to 1.0 indicates excellent class separability.[6]

---

5. Conclusion

5.1. Summary of Findings

This report has detailed a comprehensive framework for the design, implementation, and evaluation of an automated system for brain tumor classification from MRI images. The project's methodology is centered on the use of the MATLAB Deep Learning Toolbox, leveraging high-quality public datasets [3] and a rigorous preprocessing pipeline that includes mandatory skull stripping.[5]

The analysis concludes that while a custom-built CNN can provide a functional baseline, a transfer learning approach using a pre-trained ResNet architecture is capable of achieving state-of-the-art performance.[18] Benchmark results indicate that this model can achieve accuracy and F1-scores approaching 99% [6], demonstrating its high technical viability. The mathematical analysis of the CNN components confirms their suitability for this task, with convolutional layers providing hierarchical, invariant feature extraction [28], ReLU enabling the modeling of complex non-linear patterns [71], and backpropagation serving as the engine for learning.[86]

This system offers a powerful, objective, and indefatigable "second opinion" that can augment the manual, subjective process of radiological diagnosis [1], ultimately aiming to reduce diagnostic errors and improve patient outcomes.

5.2. Limitations and Future Challenges

Despite high technical accuracy on curated datasets, the transition of such a system into a real-world clinical workflow is fraught with significant challenges.[8]

1. **The "Black Box" Problem (Interpretability):** CNNs are notoriously opaque "black boxes".[102] A model that outputs "Glioma" with 99% confidence is difficult for a clinician to trust if it cannot *explain* its reasoning. Why did it choose Glioma? What features did it see? This lack of interpretability is a primary barrier to clinical adoption.[8]
2. **Data Scarcity and Generalization:** The models are only as good as the data they are trained on.[102] A model trained on a specific public dataset may fail to *generalize* when faced with real-world data from a different hospital, a 10-year-old MRI scanner, a different acquisition protocol, or a patient with a rare comorbidity.[8]
3. **Data Bias:** Deep learning models will inherit and *amplify* any biases present in the training data.[8] If a dataset is not sufficiently diverse in terms of patient age, sex, and ancestry, the resulting AI tool may be less accurate for under-represented demographic groups, thus exacerbating health inequities.
4. **Data Privacy and Security:** The solution to the generalization and bias problems is to train on massive, diverse, multi-institutional datasets. However, patient data is (and must be) protected by stringent privacy regulations (e.g., HIPAA). This creates a central paradox: the data needed to build the best tools is the most difficult to access, share, and aggregate.[8]

5.3. Future Directions in Neuro-Oncology

The work detailed in this report is a foundational step of classification. The future of AI in neuro-oncology [24] lies in moving toward more sophisticated, integrated applications:

- **Radiomics and Radiogenomics:** This involves using deep learning to extract thousands

of quantitative, sub-visual "radiomic" features from an image.[13] The next step is "radiogenomics," which correlates these image features with the tumor's genetic profile. This could lead to a "digital biopsy," where a model could predict a tumor's molecular subtype (e.g., MGMT promoter methylation status [40]) non-invasively, directly from the MRI scan.

- **Segmentation and 3D Modeling:** Moving beyond 2D classification (this project's scope) to 3D, pixel-level segmentation (as seen in the BraTS challenge [39]). A 3D model that precisely delineates the tumor, edema, and necrotic core is invaluable for surgical planning and radiation therapy.
- **Clinical Workflow Integration:** The ultimate goal is not a standalone MATLAB script but a tool that is seamlessly integrated into the radiologist's daily workflow (e.g., the Picture Archiving and Communication System, or PACS). This tool would pre-analyze scans, automatically highlight suspicious regions, provide a preliminary classification, and auto-populate parts of the radiology report, functioning as a true digital assistant that *augments*, rather than replaces, the human expert.[25]

---

6. References

1. Brain Tumor Detection Using Magnetic Resonance Imaging and ..., accessed November 12, 2025, https://www.mdpi.com/2504-2289/8/9/123
2. Item - brain tumor dataset - figshare - Figshare, accessed November 12, 2025, https://figshare.com/articles/dataset/brain_tumor_dataset/1512427
3. Brain Tumor MRI Dataset - Kaggle, accessed November 12, 2025, https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset
4. Conventional and Deep Learning Methods for Skull Stripping in Brain MRI - MDPI, accessed November 12, 2025, https://www.mdpi.com/2076-3417/10/5/1773
5. Lightweight CNN for accurate brain tumor detection from MRI with limited training data - NIH, accessed November 12, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC12426117/
6. https://gamma.app/docs/Applying-CNN-on-MRI-Images-to-Diagonise-Brain-Tumors-6t1o4vuneciy6p6
7. https://www.researchgate.net/profile/Satyabrata-Aich/publication/342048436_A_CNN_based_Approach_for_the_Detection_of_Brain_Tumor_Using_MRI_Scans/links/5edf9dc6a6fdcc4768910758/A-CNN-based-Approach-for-the-Detection-of-Brain-Tumor-Using-MRI-Scans.pdf