Tumor Detection in Brain MRI Scans using K-Nearest Neighbors

Jonathan Anguiano

August 1, 2025

Abstract

This paper documents the implementation of a K-Nearest Neighbors (KNN) classifier for distinguishing between healthy and tumor-affected brain MRI images. I trained and evaluated the model using extracted features from 3D MRI scans and integrated the system into a Django web application for live predictions and metric display.

1 Introduction

Brain tumors require early and accurate detection for effective treatment. Medical imaging, specifically Magnetic Resonance Imaging (MRI), provides non-invasive insight into brain health. In this study, I develop a machine learning classifier using the KNN algorithm to classify MRI scans into *healthy* and *tumor* categories.

2 Methodology

2.1 Data Preprocessing

The dataset consisted of pre-labeled MRI scans in `.nii.gz` format. All images were normalized and resized to a uniform shape of $64 \times 64 \times 64$ voxels before being flattened into vectors.

2.2 Model Training

Let $X \in \mathbb{R}^{n \times d}$ be the matrix of n samples each with d features, and $y \in \{0, 1\}^n$ be the corresponding labels. The KNN classifier learns by storing all training samples. During inference, a test point x_{test} is assigned a label $\hat{y}$ via:

$$\hat{y} = \text{mode}(y_i \mid x_i \in \mathcal{N}_k(x_{\text{test}}))$$

where $\mathcal{N}_k(x_{\text{test}})$ is the set of k nearest neighbors to x_{test} under the Euclidean distance.

2.3 Evaluation Metrics

The model is evaluated using:

- **Accuracy:** $\frac{\text{TP+TN}}{\text{Total}}$
- **Precision:** $\frac{\text{TP}}{\text{TP+FP}}$
- **Recall:** $\frac{\text{TP}}{\text{TP+FN}}$
- **F1 Score:** harmonic mean of precision and recall

3 Implementation

The model is trained via a Django management command and results are saved as `.joblib` files. A web interface is created using Django, allowing users to trigger training and view evaluation results in real time.

4 Results

The classifier achieved an accuracy of approximately 92.5% on test data. Precision and recall scores indicated balanced performance. **Charts and graphs will be generated and entered below.**

5 Conclusion

KNN can serve as a strong baseline for brain MRI classification tasks. Future improvements include dimensionality reduction (e.g., PCA), deeper preprocessing pipelines, and neural network baselines.

Resources

- Source code: `https://github.com/jonangui23/mri_brain_classifiers`