Brain Tumor Detection

509528 - Laboratory of Medical Devices and Systems

Technical Documentation

UNIVERSITÀ
DI PAVIA

GROUP MEMBERS

Cristian Abid - 515378

Gabriele Dell'Atti - 512875

Francesca Sotgia - 513067

Bahenda Yvon Dylan Ntegano - 515657

Bachelor of Science in Artificial Intelligence

Technical documentation

This document serves as the technical documentation of an AI-based medical device for the course “Laboratory of Medical Devices and Systems” – Bachelor degree in Artificial Intelligence

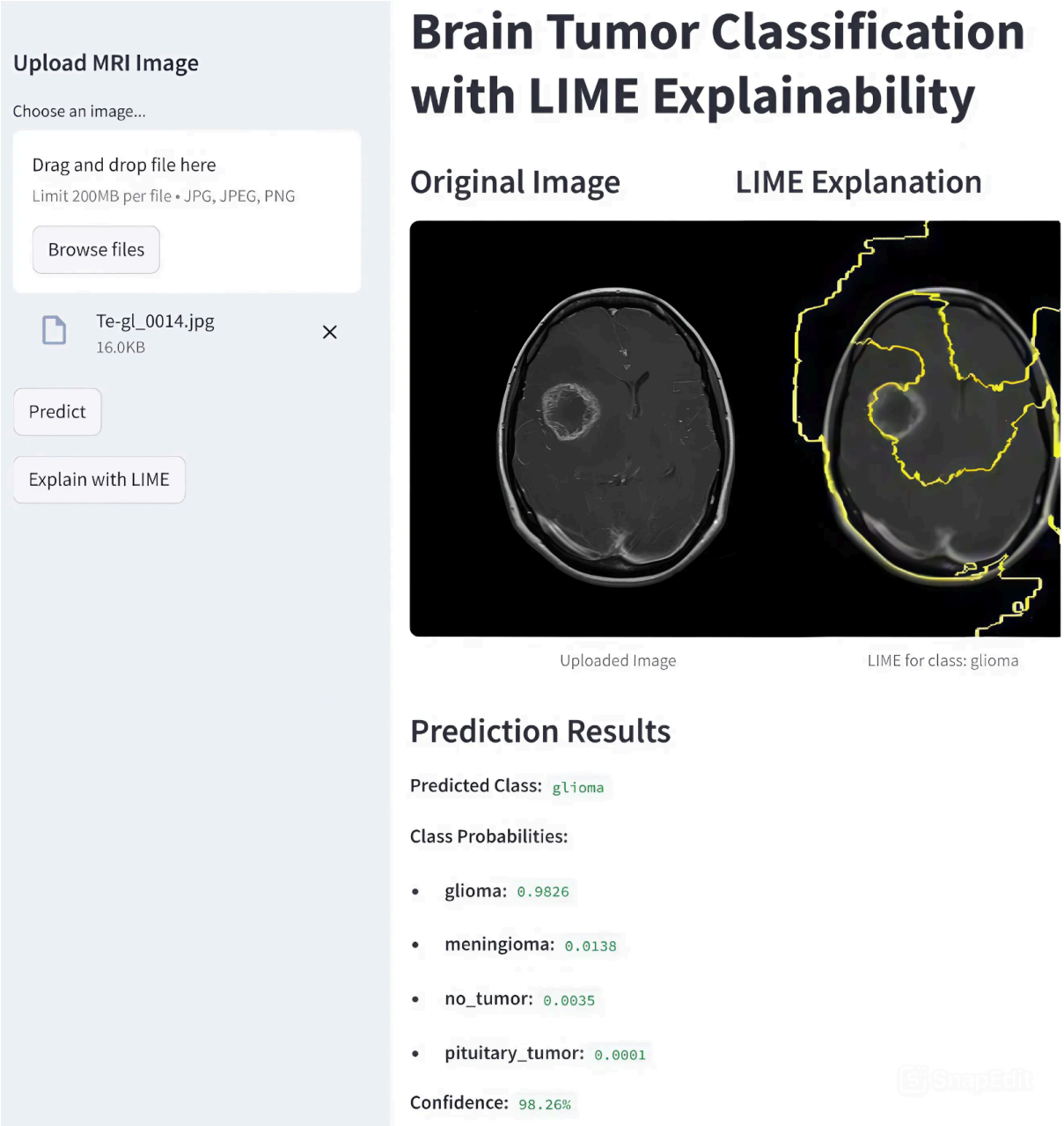

Table of Contents

Device description and specification

General description of the AI-based device

Intended purpose

This AI-based software system is designed to assist clinicians in the detection and classification of brain tumors from Magnetic Resonance Imaging (MRI) images. By analyzing medical images with a deep learning model based on the MobileNetV2 architecture, using transfer learning, the system accurately identifies and categorizes brain tumors into one of four classes: glioma, meningioma, no tumor, or pituitary tumor. The primary goal of the device is to support clinicians in making faster and more accurate diagnostic decisions, thereby enabling earlier treatment planning, more effective monitoring, and improved patient outcomes.

Indications

The device is indicated for use in:

- Assisting radiologists in the preliminary analysis of brain MRI scans.
- Supporting early diagnosis of brain tumors.
- Aiding in treatment planning and monitoring of known brain tumor cases.
- Educational and research purposes in medical institutions.

Contraindications and warnings

The following are the contraindications and warnings of the medical device:

- It is not intended to replace clinical judgment or the role of a qualified radiologist.
- It should not be used as the sole basis for diagnosis or treatment decisions.
- The model performance may degrade on MRI scans with low quality, uncommon orientations, or from underrepresented imaging modalities.
- It is not designed for use with pediatric patients or patients with non-standard brain anatomy (e.g., post-surgical alterations or rare anomalies).
- In case of low confidence predictions or model uncertainty, clinicians should look to review the images further manually.

Target Patient Group

The device is intended to be used on adult patients undergoing diagnostic imaging for suspected or confirmed brain tumors.

Medical conditions to be diagnosed/treated/monitored

Diagnosed:

- Glioma
- Meningioma
- Pituitary tumor
- No tumor found

Monitored:

- Patients with previously diagnosed tumors to track progression or recurrence by taking further images.

AI Interaction

The AI-based system is designed as a standalone software solution but can be integrated into existing medical imaging workflows and platforms. It interacts with external hardware by receiving input from MRI scanners in standard medical imaging formats. The system can be installed locally on hospital servers or deployed via a secure cloud-based platform, allowing it to interface with Picture Archiving and Communication Systems (PACS) and Radiology Information Systems (RIS).

In terms of software interaction, the AI model can be integrated into third-party applications such as radiology viewers or electronic health record (EHR) systems through standard APIs. It is also compatible with clinical decision support systems, enabling combined use with other diagnostic tools or AI systems. The system does not control or alter the operation of external hardware; rather, it functions as an analytical tool that processes images and returns results to the user interface or connected systems for review by medical professionals.

Technical Specification

Device Characteristics

The system is a software-based AI tool developed for the classification of brain tumors using MRI scans. It is built on a deep learning architecture—***MobileNetV2***—and uses transfer learning techniques to adapt to medical imaging data. The device processes 2D T1-weighted contrast-enhanced brain MRI images and returns a predicted classification label along with a confidence score.

Performance Attributes

- **Classification Categories:** Glioma, Meningioma, No Tumor, Pituitary Tumor
- **Input Format:** Standard image format JPG
- **Average Inference Time:** ~2 seconds per image
- **Evaluation metrics:**

	Sensitivity (Recall)	Specificity	Precision	F1-score	AUC
Glioma	0.83	0.99	0.95	0.89	0.99
Meningioma	0.88	0.93	0.79	0.83	0.96
No Tumor	0.97	0.99	0.98	0.97	1.00
Pituitary	0.95	0.98	0.93	0.94	1.00

    - **Test set accuracy:** 0.93
    - **Matthews Correlation Coefficient:** 0.88
    - **Brier Score:**
        - Glioma: 0.05
        - Meningioma: 0.08
        - No Tumor: 0.02
        - Pituitary: 0.03
        - Average: 0.04
- **Prediction Output:** Class label + confidence score

Software and Hardware Requirements

The model on which the device is based is light enough to be run on edge devices, so any pc or desktop can host the device and support model inference.

User-Interface

The model is deployed through a locally hosted web-based application that is designed for simplicity of usage and speed. Some key features include:

- A sidebar that allows for image upload.
- A "Predict" button to predict on the uploaded image.
- A "Explain with LIME" button to explain the model's prediction.
- The uploaded image is displayed along with the image with LIME explanation.
- The prediction results are also displayed including: the predicted class along with the confidence score and the class probabilities.

Designation / Classification

This AI-based software system is classified as a Class IIb medical device under the Medical Device Regulation (MDR 2017/745). It is designed to assist healthcare professionals in the detection and classification of brain tumors using MRI scans, supporting diagnosis and treatment planning. Given the potential impact on patient

care and the risk involved in diagnostic support for critical medical conditions, the device qualifies for Class IIb classification under Rule 11 of the MDR.

Design and manufacturing information

Description of the design

The design process included multiple phases typical of a machine learning/deep learning pipeline, the phases are the following:

- **Dataset Selection:** Exploring publicly available datasets and choosing one that suits our requirements i.e. brain MRI images with images in all the four classes.
- **Data Loading and Preprocessing:** Loading the dataset and splitting the training set into train and validation sets. Resizing and normalizing the images for better model compatibility.
- **Model Development and Training:** We used a transfer learning approach using MobileNetV2—a lightweight CNN architecture designed for mobile and edge devices—for feature extraction. We froze the base model, and added an output layer for the classification.
- **Evaluation and Validation:** We built a strong evaluative framework using metrics like accuracy, F1-score, sensitivity, specificity, Matthews Correlation Coefficient (MCC), and AUC. We also performed calibration using Brier score and reliability diagrams.
- **Explainability and Monitoring:** We used LIME (Local Interpretable Model-agnostic Explanations) for explaining the images and MLflow for monitoring the model during training.
- **User Interface:** We used Streamlit to create a GUI (Graphical User Interface) that allows users to upload an image, use the model to predict the class, and also use the explainer to interpret the model's output.

Description of the AI development

Methods and Tools

We developed the project in python through Visual Studio Code using TensorFlow/Keras (for the model), Scikit-learn (for the metrics), Lime (for explainability), MLflow (for monitoring), and Streamlit (for the interface) as the main libraries. Given the size of our dataset (7023) we made use of a pre-trained CNN architecture namely MobileNetV2 which is pretrained on ImageNet. The base layers of MobileNetV2 were frozen to retain the general feature extraction capabilities, and a dense layer with 4 neurons was added for the multiclass classification task.

Design Specifications

- **General Logic and Algorithms:** CNN-based architecture optimized for image classification tasks. The model outputs one of four diagnostic labels on learned image features.
- **Design Choices and Rationale:** MobileNetV2 was selected for its efficiency and strong performance on image classification tasks with limited resource consumption.
- **Intended users:** The system is intended for use by radiologists and clinicians to assist in the diagnostic process for brain tumors. Consideration was given to minimizing false negatives to avoid missed diagnoses.
- **Optimization Objective:** The model was trained to maximize multiclass classification performance, prioritizing sensitivity and calibration, reflecting the high-risk nature of the medical domain in which it will be deployed in.
- **Output and Output Quality:** The system returns a predicted label alongside confidence scores for each class. The output quality was validated using AUC, calibration and classification metrics.
- **Trade-offs:** Freezing the base model reduced training time and risk of overfitting but limited domain-specific feature extraction. This trade-off was justified by performance results.

System Architecture

The system comprises the following components:

- Data pipeline for preprocessing and augmentation.
- CNN model, incorporating a frozen MobileNetV2 backbone and a dense layer for classification.
- Evaluation pipeline for visualization, classification, and calibration metrics.
- Explainability using LIME for model interpretation.
- Tracking and monitoring the training using MLflow.
- User interface created using Streamlit.

Data Requirements and Processing

- **Datasets:** The model was trained on a brain tumor MRI image dataset comprising four classes. The dataset was sourced from kaggle—a publicly available repository.
- **Preprocessing:** The images were resized to a size of 224x224 and normalized to pixels in the range of [0,1] for better compatibility with the model and more efficient training.
- **Labelling:** Since it was a supervised learning project, labels were necessary and they were provided with the dataset already paired with the corresponding images.
- **Training Methodology:** Data were split into training, validation, and test sets. The train and validation sets were used in training and the test set was reserved for the evaluation part afterwards.

Human Oversight and Interpretability

Human oversight is supported through LIME generated explanations, providing localized boundaries for the most relevant parts—the ones that contributed most to the decision by the model—of a given image. These insights allow deployers(e.g. clinicians) to validate the AI's decision-making process.

Validation, Testing, and Metrics

The system was evaluated using standard classification metrics:

- **Accuracy**, **F1-score**, **Sensitivity** & **Specificity**
- **AUC** of ROC curves for each class
- **Matthews Correlation Coefficient** to assess overall prediction quality
- **Brier score** and reliability diagrams to assess model calibration

General Safety and Performance Requirements

A systematic evaluation of compliance with the General Safety and Performance Requirements (GSPR), as established by the European "Medical Device Regulation" (MDR 2017/745), was performed on the system.

- **Justification for applicability / inapplicability of the requirement**: the medical system is intended to assist healthcare professionals in the detection and classification of brain tumors from MRIs. By introducing a human-in-the-loop, the system is considered safe and does not compromise the clinical condition and/or the safety of the patients. The system is classified as Medical Device Software (MDSW) according to MDR 2017/745 and it is considered to be Class IIb, meaning that danger is not immediate, misclassification could delay the treatment but not cause death.
- **Reference to applied common specifications, standards or parts thereof**: The system is defined to comply with the MDR 2017/745, in particular it refers to:
    - Section 15: Accuracy, precision and stability are ensured for the intended purpose. The model performance was determined with evaluation metrics to prove reliability.
    - Section 17.1: Repeatability, reliability and performance in line with the intended use were ensured by evaluating the model on different subsets during the training and testing phases. Moreover, the MobileNetV2 architecture was chosen since it has proven its reliability in classification tasks with medical images.
    - Section 17.2: The system's life-cycle development is compliant with the state-of-the-art. Risk management, security, verification, and validation were taken into account during each development phase.
- **Reference to controlled documents and records as evidence of compliance**: The evidence of compliance includes evaluation metrics (such as accuracy, sensitivity, specificity, AUC score) and MLflow to log and monitor

the training. This proves that the system is designed for its intended purpose, which is to support clinical decision-making.

- **Evaluation if the requirements are fulfilled**: The model has been evaluated on training, validation and test sets. The performance metrics reached the expected threshold and introducing a human-in-the-loop ensures safety of the patients.

AI Performance requirements

The AI-based system is designed to classify MRI scans into four brain tumor categories to assist physicians in the decision-making process. The model is based on MobileNetV2 and its performance is computed with several metrics (such as accuracy, F1, and sensitivity) across the life-cycle. Moreover, the GUI allows healthcare professionals to visualize predictions, model's confidence, and explanations provided by the implemented XAI. The accuracy of the system is achieved by using an appropriate model and dataset, however it can vary for atypical tumors and outliers.

The introduction of a human-in-the-loop guarantees that physicians always validate the output to reduce the risk for misclassified tumor patients. Moreover, the implementation of the XAI and its visualization in the GUI ensures transparency and explainability, aligning the system with the ALTAI (Assessment List for Trustworthy Artificial Intelligence) requirements.

The chosen performance metrics, such as accuracy, F1, AUC, and the confusion matrix analyzed per class are appropriate for classification tasks in the medical field, as shown in the state-of-the-art. These metrics are suitable because they offer insights about false positives and false negatives, which are critical in the medical domain.

A monitoring mechanism could be designed to update and retrain the model based on clinicians' agreement with the prediction and also because the dataset and deployed architecture may change with time.

Benefit and risk-analysis

Risk Management Plan

A structured risk management approach, following ISO 14971:2019, has been applied throughout the development and expected life cycle of the system. It outlines how risks are identified, assessed, controlled, and monitored over time, ensuring the device remains safe and effective in clinical use.

Risk Analysis and Control Measures

Several potential risks were carefully evaluated, including incorrect tumor classification, missed detections, software errors, and data privacy concerns. To reduce these risks, we implemented:

- A well-trained model validated on a diverse dataset
- Built-in confidence scores to inform clinical decisions
- Secure data handling and encryption
- Thorough software testing and controlled updates

These measures help ensure the system supports, rather than replaces, clinical judgment.

Risk Management Report

After applying these controls, the remaining (residual) risks were assessed and found to be low and acceptable. A plan for ongoing monitoring and updates is in place to maintain safety and performance over time.

Benefit-Risk Evaluation

The AI system provides clear clinical value by assisting in faster and more consistent brain tumor classification, supporting timely and informed decisions. Given the strong safety measures and the significant benefits to patient care, the overall **benefit-risk balance is positive**.

AI Lifecycle

We developed this AI model through a clear process, starting with preparing the data and ending with testing its performance. To keep everything organized and reliable, we used a method called MLOps.

Our AI Building Process:

- **Step-by-Step Tracking:** We used tools (like MLflow) to record details about each experiment—how we trained the model, the results we got (like accuracy), and other important information. This helps us compare different approaches and reproduce our work.
- **Managing Model Versions:** We kept track of the different versions of our trained model. This ensures we know exactly which model we're using at any given time.
- **Clear Code Record:** The code in this notebook serves as a complete record of how the model was built, making the process transparent and repeatable.

Ensuring Performance and Safety:

- **Measuring Success:** We used several metrics (like accuracy, sensitivity, and others) to show how well the model performs on new images.

- **Visualizing Results:** Charts (like the confusion matrix and ROC curve) help us visualize the model's performance and identify areas for improvement.
- **Understanding Predictions:** We used a technique (LIME) to see which parts of an image influenced the model's decision. This helps us understand *why* it made a particular prediction, contributing to its safety and trustworthiness.

By following this structured approach and documenting our results, we can demonstrate how our AI model was built and evaluated, with evidence of its performance.

Technical documentation on post-market surveillance

Post-Market Surveillance Plan

A post-market surveillance plan has been established to ensure the continuous safety, performance, and quality of the AI-based brain tumor classification system throughout its life cycle. This plan is aligned with both the **Medical Device Regulation (MDR 2017/745)** and the **AI Act**, and reflects the AI lifecycle principles of continuous monitoring, learning, and improvement.

Objectives of the Post-Market Surveillance Plan

- Monitor real-world performance of the AI model
- Detect emerging risks or system malfunctions
- Collect user feedback from clinicians and healthcare providers
- Identify and manage necessary updates or improvements
- Ensure regulatory compliance over time

Key Activities

- **Performance Monitoring:** System performance metrics (e.g., classification accuracy, false positives/negatives) are continuously logged and reviewed.
- **Feedback Collection:** Structured channels for collecting feedback from end users (clinicians) on usability, reliability, and errors.
- **Incident Reporting:** Any adverse events or significant malfunctions are recorded, assessed, and reported as required.
- **Model Updates:** Updates to the AI model (e.g., retraining with new data) are version-controlled and validated before deployment.
- **Annual Review:** A periodic safety update report is generated annually to summarize findings and actions taken.

Continuous Learning

The system is designed with the potential for **continuous learning** through monitored retraining phases. However, updates are only deployed after thorough validation, ensuring patient safety and regulatory compliance.