TeqFest Hackathon Project

MRI Scans' Brain Tumor Detection Algorithm

Team members
Selsabeel Asim Ali
Basma Mahmoud Hashem
Roaa Fathi Nada
Nourhan Mahmoud Eldesouky

Objectives and Significance:

Our project aims to contribute to the Sustainable Development Goal (SDG) 3: "Ensure healthy lives and promote well-being for all at all ages." We chose this goal since we felt that it is easily the most important goal, without it we cannot achieve other SDGs. SDG 3 focuses on preventing unnecessary human suffering around the world, working on improving the overall health of populations all around the world. SDG also emphasizes the importance of increasing investment in research & development, health financing, and health risk reduction and management. The current trend in health research & development is finding use cases for the ever-changing artificial intelligence algorithms.

We chose to specifically focus on how we can use AI to address the impact of diseases on global health, specifically cancer. Millions of people are diagnosed with cancer around the world every year, and more than half of the patients eventually die from it. This number is astounding and there is great potential in working to decrease the mortality rate of this disease through early detection. That's why we chose to work on a brain tumor MRI scans dataset where early detection will be most crucial in saving the patients. We used artificial intelligence, computer vision, and deep neural networks, to tackle these challenges.

Description of our Model:

Our model uses computer vision techniques and deep neural networks to address various health-related challenges, particularly in the context of tumor detection from medical images. The model's architecture consists of the following components:

- Data Preprocessing: Using the libraries TensorFlow and Pandas which are vital for any data science/AI project, to extract relevant features and prepare them for input into our neural network model. Simple tasks like resizing the data and ensuring it is in the correct form for easy usage by our CNN model.
- Convolutional Neural Network (CNN): The model's most important algorithm and the main building block is CNN, a deep learning architecture known for working well on analyzing unstructured data like MRI scans that would need computer vision. The model uses TensorFlow's built-in functions to discover features in the input images.
- Transfer Learning: Due to the model's relatively small training dataset size (not that small but not large enough to train enough on its own) we made use of previously built models that were built on known large datasets. We used ResNet50. We then used it to fine-tune our model's parameters to make the model better at detecting the features of a brain tumor in the brain scan.
- Classification: The model is trained for binary classification, specifically to classify whether an input image contains a tumor or not. Using two datasets representing tumor and non-tumor images, the model is trained to predict a binary value (yes, contains a tumor or no) based on the features extracted by the CNN architecture.

Limitations of the Model:

- **Long Training Time Combined with Limited Hackathon Time:** This was easily our biggest limitation, as we wanted to try many different things to test if they'd improve the model's accuracy however due to the long training time of our model it was difficult to test different things in our algorithm like tuning hyperparameters and changing the CNN layer structure to optimize it.
- **Limited Dataset Size**: The model's performance is hindered by the size and diversity of the dataset used for training. A larger and more varied dataset could potentially improve its generalization. It also would've benefited from using 3D images of brain scans instead of 2D image scans, but we couldn't find any such public dataset.
- **Lack of Sufficient Data Pre-processing**: It is indeed a very important step in any Machine Learning algorithm, however due to the limited time of the hackathon we were not able to add as much data pre-processing in our model as we would have liked.
- **Avoided Using Data Augmentation:** We were advised not to use data augmentation for a hackathon as it may overinflate our accuracy and giving our dataset a bigger proportion of the accuracy than it really is. If it weren't a hackathon, the model could've greatly improved from using data augmentation to increase the dataset size and improve the computer vision algorithm implemented by making the computer better at detecting the same features even if the position or angle changes.

• **Simplistic Architecture**: The model's architecture, although effective, is relatively simple. More complex architectures or ensembling techniques could potentially enhance its performance further.
• **Fixed Image Size**: The model operates on images resized to 128x128 pixels. This fixed size may limit its ability to capture fine details present in higher-resolution images.
• **Binary Classification**: The model is designed for binary classification (tumor vs. no tumor). Extending it to handle multiple tumor types could greatly enhance our model.
• **Limited Interpretability**: Deep learning models like this one often lack interpretability, making it challenging to understand the reasoning behind specific predictions. It feels almost like a black box that changes the weight parameters but wouldn't be of use for a field specialist to understand what the algorithm uses to find features that distinguishes a brain tumor in a MRI scan.

Comparison with Existing Algorithms & Approaches:

Before deciding to go through with this project, we tried some of the already available algorithms avalaible on Kaggle and tested them on different datasets. Their issue was that they overfitted for performance on the dataset they trained on, however, naturally, they didn't perform as well on benchmark datasets despite them being visually very similar to the dataset they initially trained their CNN model on. There were even notebooks with 100% accuracy that performed very poorly when validating on other similar datasets of MRI brain scans. The real struggle of the project we made was to build an algorithm that performs well even with new MRI scans it hadn't seen before from new datasets.

The current algorithms developed are limited by the small datasets available, as is the case for ours' as well. We would greatly benefit from gaining access to a larger dataset.

Moreover, according to a research paper by the University of Auckland, despite there being many different approaches that were explored by scientists for MRI scans analysis, ML-based methods have been proven to perform significantly better in this task than statistical methods. Transfer learning by using prebuilt models and existing computer vision models greatly improve the accuracy of these algorithms. However, there isn't any proof that we have to use ML models for this task.

Source: University of Auckland. (Year). Exploring machine learning methods for MRI scan analysis. Journal of Medical Imaging, 10(4), 397. https://www.mdpi.com/2306-5354/10/4/397