**Name:** Barath Nochipalayam Subramaniam

**Student ID**: 22085249

**Project and Data Management (PDM) Plan: Brain Tumour Classification Through MRI Images**

**1. Project Overview**

**Project Title:**

Brain Tumour Classification Through MRI Images Using Convolutional Neural Networks (CNNs)

**Summary of the Project Topic and Background:**

Brain tumours are pathological lesions of brain tissue or central spine origin that may interfere with normal brain functions. It widely known that the early diagnosis of brain tumours dramatically improves the treatments available and chances of the patient’s survival. MRI (Magnetic Resonance Imaging) is the best equipment in the diagnosis of brain tumours since they provide detailed structures in the brain. However, MRI scans analysis is a time consuming and prone to errors due to the interference of human beings.

In recent years, the degrees of achievements in the medical image classification using deep learning techniques especially CNNs are remarkably promising. CNNs are very good for image-related problems, for instance, recognizing features in MRIs that are related to particular tumour stage. This project is employed CNNs for the classification of MRI images of brain tumours into glioma, meningioma, pituitary vi tumours, and non-tumour cases. Kaggle dataset will be utilized for model training and evaluation.

**Aim:**

To classify brain tumour MRI images using a machine learning model, specifically a Convolutional Neural Network (CNN), to improve diagnostic accuracy.

**Project Objectives:**

1. Preprocess the MRI image data for training the model.
2. Train a CNN model to classify tumour types.
3. Evaluate the model’s performance using accuracy, precision, and recall.
4. Explore possible improvements and future work.

**Research Question:**

How can convolutional neural networks (CNNs) improve the speed and accuracy of brain tumour classification from MRI images compared to traditional manual diagnostic methods?

**Reference list**

Vankdothu, R. and Hameed, M.A. (2022). Brain Tumor MRI Images Identification and Classification Based on the Recurrent Convolutional Neural Network. *Measurement: Sensors*, 24, p.100412. doi:https://doi.org/10.1016/j.measen.2022.100412

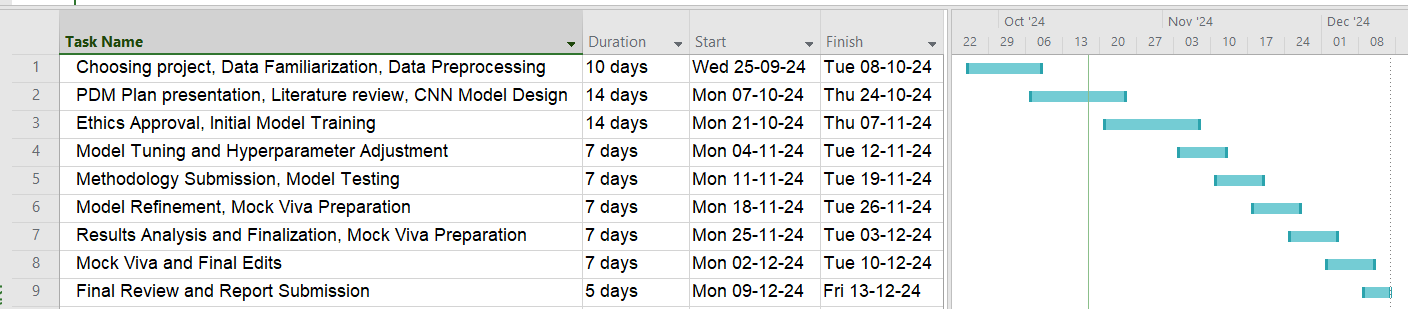
Hu, H., Li, X., Yao, W. and Yao, Z. (2021). *Brain Tumor Diagnose Applying CNN through MRI*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICAICE54393.2021.00090.

Ruqsar Zaitoon and Syed, H. (2024). A Comprehensive Review: Tumour Recognition, Segmentation and Classification in Brain MRI. *IEEE Xplore*, [online] pp.865–870. doi:https://doi.org/10.1109/icoeca62351.2024.00152.

**2. Task List**

|  |  |
| --- | --- |
| **S.No** | **Task** |
| 1 | Choosing project, Data Familiarization, Data Preprocessing |
| 2 | PDM Plan presentation, Literature review, CNN Model Design |
| 3 | Ethics Approval, Initial Model Training |
| 4 | Model Tuning and Hyperparameter Adjustment |
| 5 | Methodology Submission, Model Testing |
| 6 | Model Refinement, Mock Viva Preparation |
| 7 | Results Analysis and Finalization, Mock Viva Preparation |
| 8 | Mock Viva and Final Edits |
| 9 | Final Review and Report Submission |

**Project Timeline**



**3. Data Management Plan**

**Overview of the Dataset:**

Overview of the Dataset: For this project, Kaggle Brain Tumour Classification dataset will be used. It contains a large set of MRI images labelled into four categories: glioma, meningioma, pituitary tumour, no tumour. The dataset includes approximately 18279 images, and those images could be of almost any size and resolution.

**Source:** <https://www.kaggle.com/datasets/rishiksaisanthosh/brain-tumour-classification/code>

**Data Format:** Images are in JPEG format with corresponding labels provided in a CSV file.

**Data Collection:** This is the reason why; the dataset is already available and will be directly downloaded from Kaggle. This data saves a lot of time and therefore effort since one does not have to go out of their way to colleting datas through other sources.

**Metadata:** The source contains MRI, which is further categorized based on the nature and type of tumour, viz glioma, meningioma and pituitary tumours, size of the images and dimensions of images. Data storage will be in the region of about 500mb space depending on the number of augmented images.

**Data Preprocessing:** Preprocessing steps will include resizing all images to a standard size to maintain consistency across the dataset. Normalization of pixel values will be performed to scale the image values between 0 and 1. Data augmentation (rotation, flipping, zooming) will be applied to avoid overfitting due to the relatively small size of the dataset.

**Document Control:** GitHub will be the primary version control system for tracking the progress of the code and model development. Regular weekly commits will be made, and clear documentation of all changes to the code and data will be maintained in the repository.

GitHub: <https://github.com/Barathnsj/rain-Tumour-Classification-Through-MRI-Images>

**Backup and Security:** Data and model files will be stored both locally on the project lead's machine and backed up in cloud storage GitHub. Patient anonymity is already ensured as the Kaggle dataset is anonymized, but further security measures will be applied to secure the data, including encryption of sensitive files.

**4. Ethical Requirements**

**GDPR Compliance**

The dataset contains no personally identifiable information (PII), ensuring compliance with the EU's General Data Protection Regulation (GDPR). Since no data identifies individuals, the dataset is anonymized, mitigating risks of data misuse. This safeguard protects privacy, a key factor in ethical medical research.

**UH Ethical Policies**

The project follows University of Hertfordshire (UH) ethical requirements. Medical research is notoriously plagued by patient data concerns. The Kaggle dataset is deidentified and does not directly use patient data, therefore it meets UH institutional ethical requirements. The ethics approval process will occur in project weeks 5-6.

**Permission to Use Data**

The dataset for this project is obtained from Kaggle website and the permission to use it for this learning purpose has been given under the terms of use. This availability of the maintained public dataset is a legal key to ensuring that the dataset available for this research project is not legally restricted for the purpose of the project’s study.

**Data Collection Ethics**

There is information that the Kaggle dataset was received from a third party and who collected the data in such a way to make correct ethical decisions. In order to mitigate risk of breaching patient privacy the developers of the dataset made efforts to strip out the personal identifying information (PII) leaving the data in the dataset de-identified for research use.