Plan Overview

A Data Management Plan created using DMPTool

Title: Brain Tumor Classification using Transfer learning

Creator: Mohammed Seid

Affiliation: University of Gondar (uog.edu.et)

Contributor: Assefa chekole

Funder: Digital Curation Centre (dcc.ac.uk)

Template: Digital Curation Centre

Project abstract:

Brain tumor is an abnormal mass of tissue in which cells grow and multiply uncontrollably and caused to one of the most dangerous cancer types in the world called Brain cancer, so thousands of people are suffering from malignant brain tumors leading to a very short expected life if diagnosed at a higher grade Depending on the level of cancer, early diagnosis and grading is a very critical step after detecting the tumor to achieve an effective treatment plan,. However, thousands of scans must be studied in order to classify tumor types with high accuracy. Deep learning models and computer vision can handle that amount of data, and they can present results with high accuracy, but they may take large computational time ,So the aims of this project is to classify MRI o four different tumor classes, one normal and three different abnormal brain tumor classes using a model from pre trained Convolutional Neural Network (CNN) called ResNet50V2 and VGG-19 by using transfer learning so solve above problem , and used The preferred image to detect brain tumors called Magnetic Resonance Imaging (MRI), The research methodology involves collecting a large dataset of MRI images of brain tumors, preprocessing the data, training and validating the CNN model from pre trained , and evaluating its performance using various metrics. The expected outcomes of this research include the development of a highly accurate and efficient brain tumor classification model that can assist clinicians in diagnosing brain tumors and improving patient outcomes. The research findings will also contribute to the growing field of deep learning-based medical image analysis and provide insights into the potential of this approach for other medical conditions

Start date: 07-03-2023

End date: 08-03-2023

Last modified: 06-17-2023

Brain Tumor Classification using Transfer learning

Data Collection

What data will you collect or create?

MRI Image Dataset: Gather a large dataset of MRI images containing both healthy and tumor-affected brain scans. These images should be labeled with their respective classes (e.g., healthy, meningioma, glioma, pituitary tumor, etc.).

How will the data be collected or created?

the MRI data set can be collecting by visiting different type of publickly available data set such as

The Brain Tumor Dataset available from Kaggle (https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection)
The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) dataset (http://braintumorsegmentation.org/)

Documentation and Metadata

What documentation and metadata will accompany the data?

This is a mock project on this topic, i would not actually collect or generate any real data. However, i would still need to specify what documentation and metadata will accompany the data in such a project to demonstrate my data management planning

- **Imaging metadata**: such as image modality The type of imaging technology used to acquire the image, such as magnetic resonance imaging (MRI)
- **Quality control measures**: Information about any quality control measures taken during image acquisition or processing, such as image artifact removal or normalization
- **Data dictionaries**: A data dictionary for the research data may include information about the features extracted from the MRI images, such as shape, texture, and intensity
- **Readme files**: A readme file may provide information about the structure and format of the research data, as well as any special instructions or requirements for using the data.
- **Codebooks**: A codebook may describe the coding scheme used for the machine learning algorithms and provide information about how variables were coded and how missing data were handled ***example*** py (python)file
- **Model metadata**: Metadata for the machine learning models used for brain tumor classification may include information about the model architecture, training data, hyperparameters, and evaluation metrics

Ethics and Legal Compliance

How will you manage any ethical issues?

Managing ethical issues is an important consideration for any research project, including this mock project focused on brain tumor classification using transfer learning. Here are some steps that could be taken to manage ethical issues in this project:

- **By Anonymizing data**: To protect the privacy of the participants, it may be necessary to remove any identifying information from the MRI images and associated metadata.
- **By Ensuring data security**: It is important to ensure that the research data is stored securely and only accessible to authorized personnel. This may involve using encryption, firewalls, and other security measures to protect the data.
- **By Following ethical guidelines for machine learning**: Machine learning algorithms can raise ethical concerns around issues such as bias, fairness, and transparency. Researchers should follow ethical guidelines for machine learning, such as those established by the Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) community, to ensure that their models are developed and used ethically.
- **By proper Data sharing Agreemen**t: If the data is shared with other researchers, it is important to have a data sharing agreement in place that outlines how the data can be used and protected

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

Managing copyright and intellectual property rights (IP/IPR) issues is an important consideration for any research project, including a mock project focused on brain tumor classification using transfer learning. Here are some steps that could be taken to manage IP/IPR issues in this project

- **Using open-source materials**: Whenever possible, use open-source datasets, models, and algorithms that are available under permissive licenses such as MIT or Apache 2.0,This ensures that there are no copyright or IP/IPR issues associated with the materials you are using.
- **Giving credit**: Whenever i use someone else's work, i should give credit to the original authors or owners. This helps to acknowledge their contribution and avoid any potential legal issues.
- **Consult a legal expert**: If i'm not shure about the copyright or IP/IPR status of any material, i will consult a legal expert who can advise you on the best course of action.

Storage and Backup

How will the data be stored and backed up during the research?

Storing and backing up data is an important consideration for any research project,in this mock project i will create best storage place and backup for my project by taking some steps that are listed below

- Choosing a secure storage solution
- Establishing a file organization system
- Backing up data regularly
- Maintaining data security

How will you manage access and security?

To manage access and security during the mock research project on brain tumor classification using transfer learning, i will consider the following measures

- **Implement access controls**:should implement access controls to ensure that only authorized personnel have access to the research data. This may involve using password-protected user accounts or other authentication mechanisms
- **Secure data transfer**: i should use secure data transfer mechanisms to transfer data between locations or to other researchers
- **Train personnel**: Researchers should provide training to personnel on access and security policies and procedures to ensure that they understand their roles and responsibilities in protecting the research data

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

During the mock research of brain tumor classification using MRI images, there will be several types of data that have long-term value and should be retained, shared, and/or preserved. These include:

- **Processed data**: The processed data that was used to train the machine learning models, such as feature vectors or pre-processed images, may also have long-term value and should be retained. This data can be used to compare different machine learning models or to develop new techniques for feature extraction.
- **Machine learning models**: The machine learning models that were developed and trained during the mock research have long-term value and should be preserved. These models can be used to classify new MRI images and can serve as a benchmark for future research.
- **Documentation**: The documentation associated with the mock research project, such as data dictionaries, codebooks, and readme files, should be preserved and shared

What is the long-term preservation plan for the dataset?

Data Sharing

How will you share the data?

Sharing the data after developing the mock research of brain tumor classification using MRI images is an important step in making the research accessible to other researchers and enabling future research and collaboration. Here are some steps that could be taken to share the data

- Choosing a data sharing platform: i should choose a data sharing platform that is appropriate for their needs and meets the requirements for data sharing. This may involve using a general-purpose platform, such as Zenodo or Figshare, or a domain-specific platform, such as the Cancer Imaging Archive (TCIA)

Are any restrictions on data sharing required?

Question not answered.

Responsibilities and Resources

Who will be responsible for data management?

For developing the mock research of brain tumor classification using MRI images, the responsibility for data management may fall on me (Researcher)

- **Research** : The researcher as a whole may be responsible for ensuring that the data is managed appropriately. This may involve adhering to data management policies and procedures, properly documenting the data, and ensuring that the data is accessible to other researchers

What resources will you require to deliver your plan?

the research will require various resources such as

- **MRI data**: Access to MRI images of brain tumors is essential for developing the machine learning models for the classification task
- **Computing resources**: Developing and training machine learning models requires significant computing resources such as high performance computer
- **Software tools**: i will need access to software tools for image processing, machine learning, and data analysis. This may include tools such as Python, TensorFlow, and other software packages for image analysis and machine learning
- **Personnel**: The research team will require personnel with expertise in medical imaging, machine learning, and data analysis. This may include researchers, data analysts, and software developers

---

Planned Research Outputs

Model representation - "Brain tumor classification using Transfer learning"

some potential output could be made from the research:

- **Transfer Learning Model**: This research output type could include the trained model architecture and its weights, hyperparameters, and other related information used for the brain tumor classification task. This can help ensure that the model is reproducible and can be used by other researchers for similar tasks.
- **Preprocessing Steps**: This research output type could include the preprocessing steps applied to the MRI images before feeding them into the transfer learning model. This information can be important for other researchers who want to replicate the study or apply the same preprocessing steps to their own research
- **Evaluation Metrics**: This research output type could include the evaluation metrics used to assess the performance of the transfer learning model, such as accuracy, F1-score, precision, and recall. This information can be useful for comparing the performance of different models and for assessing the reliability of the results.
- **Visualization of Results**: This research output type could include visualizations of the classification results, such as confusion matrices and heat maps, which can help to demonstrate the performance of the transfer learning model and highlight areas for improvement.
- **Code**: This research output type could include the source code used to train and evaluate the transfer learning model. This can help other researchers to replicate the study or use the code as a starting point for their own work

---

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Brain tumor classification using Transfer learning	Model representation	2023-08-31	Open	CiteSeerX	150 MB	Creative Commons Attribution Non Commercial No Derivatives 4.0 International	DataCite Metadata Schema	No	No