Brain Tumor Detection System

#1 View of Progress as of 2/4/2024

Abdullah Alshamrani
Saint Joseph’s University

- Weeks 1-2: Project start point and Data Collection
- Tasks:

Set up the project repository.
Collect and download the brain tumor dataset from Kaggle.
Upload the dataset to Google Colab for initial exploration.
Data Visualization and understanding –..--..--..--..
Begin the exploratory data analysis (EDA) to understand the dataset's characteristics.

What I have done:

1. I Set up the project repository.
2. Collected and downloaded the brain tumor dataset from Kaggle
3. Uploaded the dataset to Google Collab for initial exploration
4. Data Visualization and understanding is still under the process

What I have done:

1. **I Set up the project repository:** https://github.com/Abdullah-Alshamrani/Brain-Tumor-Detection-System
2. Collected and downloaded the brain tumor dataset from Kaggle
3. Uploaded the dataset to Google Collab for initial exploration
4. Data Visualization and understanding is still under the process

Abdullah-Alshamrani Update README.md eee94f7 · 2 minutes ago 4 Commits

README.md Update README.md 2 minutes ago

README

Brain-Tumor-Detection-System

This project aids doctors in diagnosing whether a patient has a brain tumor and identifying its stage.

Detailed Project Description:

This project should present a view of the brain and its condition regarding tumors. A medical image will be uploaded to a web page, and it should output the brain tumor classifications, whether the brain has tumors or not, this will be performed by using traditional machine learning methods and deep learning models, and depending on the result, a simple advice will be shown to the user by using ChatGPT API. Machine learning will be trained through a provided publicly available dataset through Kaggle [1].

Detailed Outline:

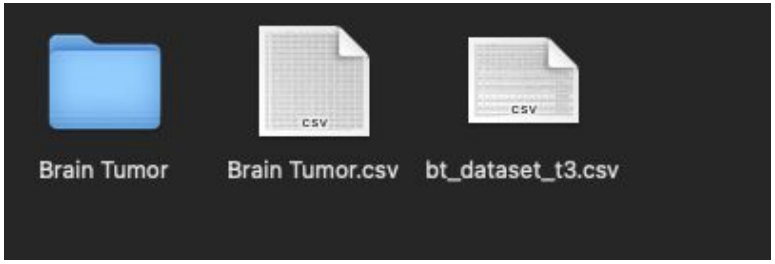
1. Using Google Colab data will be entered to introduce exploratory data analysis (EDA).
2. Data preprocessing, comparison of traditional machine learning and deep learning models.
3. Focus on convolutional neural networks by using Python to choose the most optimal machine learning model.
4. A design of the interface should introduce the user to a web page that asks the user to input a medical image of the brain.
5. It should perform with the data I have evaluated to give an output of the brain tumor detection on another web page with an advice regarding the result that is provided by ChatGPT using ChatGPT API that will be connected to my program.

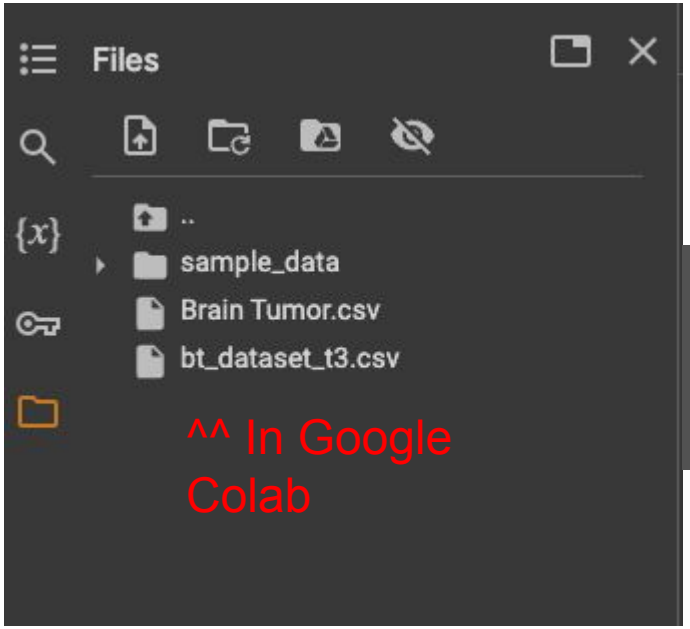
Reference:

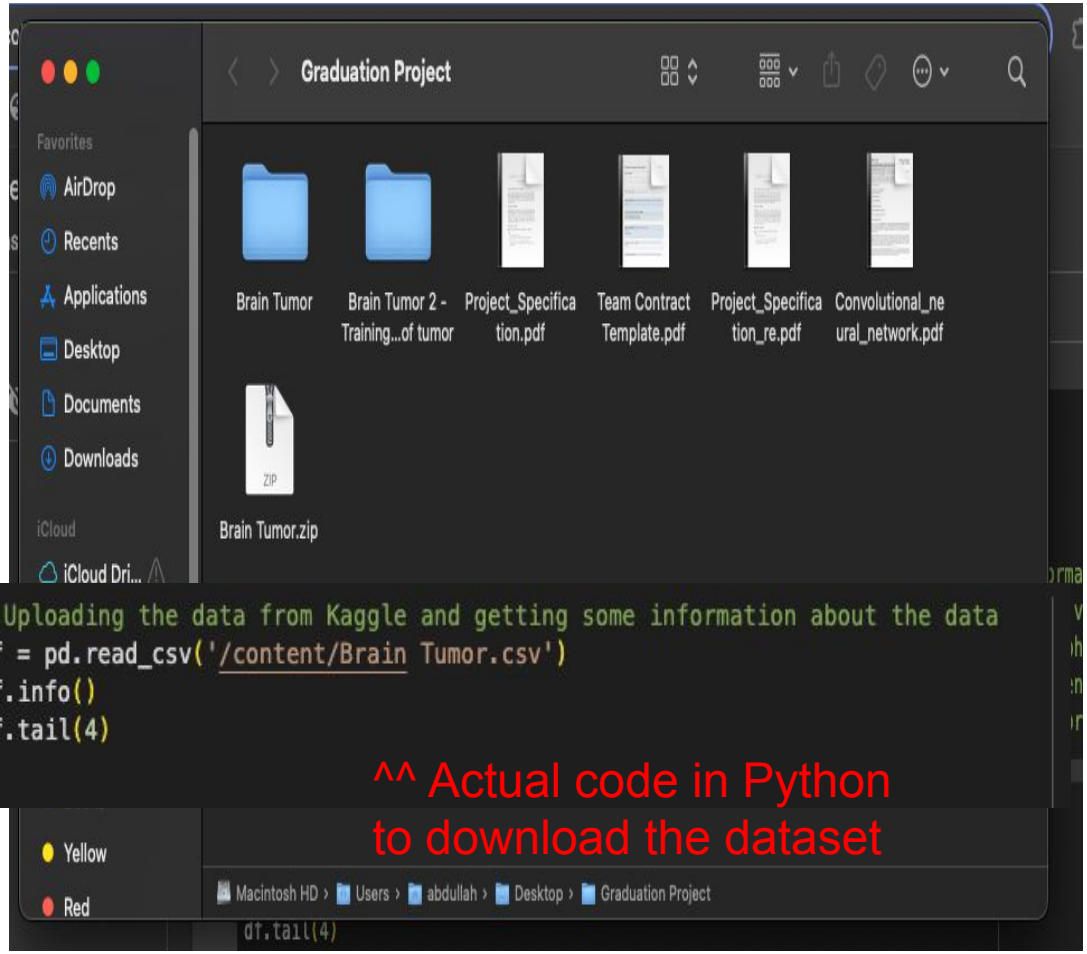
1. https://www.kaggle.com/datasets/jakeshbohaju/brain-tumor/data

What I have done:

1. **I Set up the project repository:** https://github.com/Abdullah-Alshamrani/Brain-Tumor-Detection-System
2. **Collected and downloaded the brain tumor dataset from Kaggle**
3. Uploaded the dataset to Google Collab for initial exploration
4. Data Visualization and understanding is still under the process


Brain Tumor
Brain Tumor.csv
bt_dataset_t3.csv


Files
..
sample_data
Brain Tumor.csv
bt_dataset_t3.csv
^^ In Google Colab


Graduation Project
Favorites
AirDrop
Recents
Applications
Desktop
Documents
Downloads
iCloud
iCloud Dri...
Brain Tumor
Brain Tumor 2 - Training...of tumor
Project_Specification.pdf
Team Contract Template.pdf
Project_Specification_re.pdf
Convolutional_neural_network.pdf
Brain Tumor.zip
# Uploading the data from Kaggle and getting some information about the data
df = pd.read_csv('/content/Brain Tumor.csv')
df.info()
df.tail(4)
^^ Actual code in Python to download the dataset
Yellow
Red
Macintosh HD › Users › abdullah › Desktop › Graduation Project
df.tail(4)

What I have done:

1. **I Set up the project repository:** https://github.com/Abdullah-Alshamrani/Brain-Tumor-Detection-System
2. **Collected and downloaded the brain tumor dataset from Kaggle**
3. **Uploaded the dataset to Google Collab for initial exploration**
4. Data Visualization and understanding is still under the process

[14]
```
# Loading required libraries for the project

import numpy as np
import pandas as pd
import seaborn as sns #provides a high-level interface for drawing attractive and informative statistical graphics.
import matplotlib.pyplot as plt #library for creating static, animated, and interactive visualizations in Python.
import plotly.express as px #create beautifully interactive and publication-quality graphs and charts.
import graphviz #open-source graph visualization software that provides a way to represent structural information as diagrams of abstract graphs and networks.
import sklearn as sk #open-source Python library for machine learning, offering tools for data mining and analysis.  It supports various machine-learning models for classificat
```

Required Libraries

```
# Uploading the data from Kaggle and getting some information about the data
df = pd.read_csv('/content/Brain Tumor.csv')
df.info()
df.tail(4)
```

← Actual code in Python to download the dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3762 entries, 0 to 3761
Data columns (total 15 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Image               3762 non-null   object
 1   Class               3762 non-null   int64
 2   Mean                3762 non-null   float64
 3   Variance            3762 non-null   float64
 4   Standard Deviation  3762 non-null   float64
 5   Entropy             3762 non-null   float64
 6   Skewness            3762 non-null   float64
 7   Kurtosis            3762 non-null   float64
 8   Contrast            3762 non-null   float64
 9   Energy              3762 non-null   float64
 10  ASM                 3762 non-null   float64
 11  Homogeneity         3762 non-null   float64
 12  Dissimilarity       3762 non-null   float64
 13  Correlation         3762 non-null   float64
 14  Coarseness          3762 non-null   float64
dtypes: float64(13), int64(1), object(1)
memory usage: 441.0+ KB
```

Initial exploration for the dataset preformed

	Image	Class	Mean	Variance	Standard Deviation	Entropy	Skewness	Kurtosis	Contrast	Energy	ASM	Homogeneity	Dissimilarity	Correlation	Coarseness
3758	Image3759	0	20.435349	1227.151440	35.030721	0.066763	2.144625	4.882034	161.158675	0.225931	0.051045	0.502712	5.083126	0.952749	7.458341e-155
3759	Image3760	0	18.011520	1151.582765	33.934978	0.068396	2.308349	5.579498	167.130118	0.228930	0.052409	0.492269	5.103700	0.952181	7.458341e-155
3760	Image3761	0	13.330429	945.732779	30.752769	0.087872	2.732822	7.757570	223.812932	0.261527	0.068397	0.480064	6.439784	0.940898	7.458341e-155
3761	Image3762	0	6.110138	480.884025	21.929068	0.118171	4.110669	17.538826	239.251388	0.306224	0.093773	0.494333	6.787329	0.938731	7.458341e-155

Under the process to be completed by 2/12/2024

1. Data Visualization and understanding
2. Exploratory data analysis (EDA) to understand the dataset's characteristics.

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods. It helps determine how best to manipulate data sources to get the answers you need, making it easier for data scientists to discover patterns, spot anomalies, test a hypothesis, or check assumptions.

Reference

1. **"Exploratory Data Analysis." IBM, n.d., https://www.ibm.com/topics/exploratory-data-analysis.**