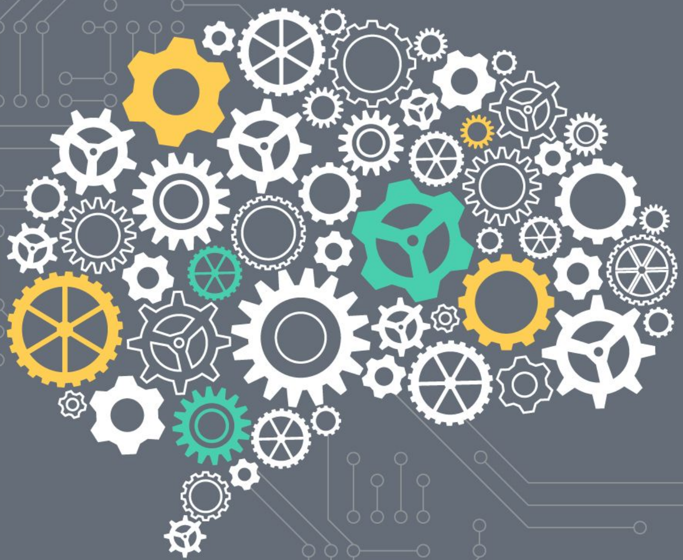


Brain Tumor Classification

MATH7375, Northeastern University Spring 2023
Topological Methods for the Analysis of Data

Josue Antonio and Olakunle Abawonse

Project Introduction

- **Goal:** To train a deep neural network (DNN) model for classification of brain tumor MRI scans.
- **Data Source:** Kaggle
- **Application:** When provided with an unlabeled brain MRI scan, the model will show us an accurate prediction of the tumor type

Methodology

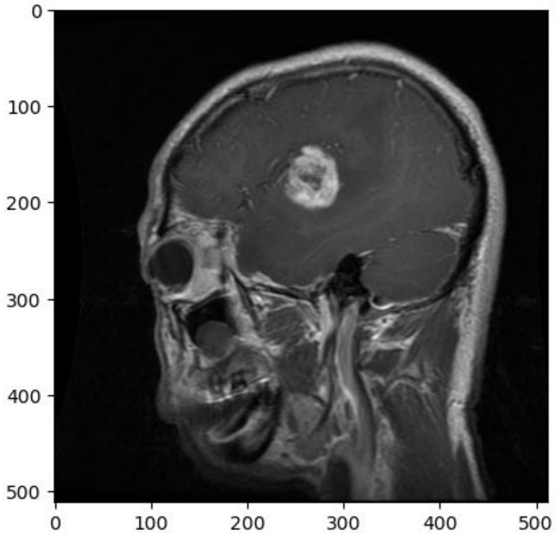
1. Given an image apply the sublevel set filtration
2. Divide each image into slices
   a. For each MRI image divide the image into horizontal and vertical slices of equal size
3. Compute corresponding persistence diagrams in 0 and 1-dimensional homologies
4. Compute corresponding persistence landscapes in 1-dimensional homology
5. Compute mean persistent landscape for each MRI image
6. Train a MLPClassifier DNN using the mean landscapes as the training input

Sublevel Filtration

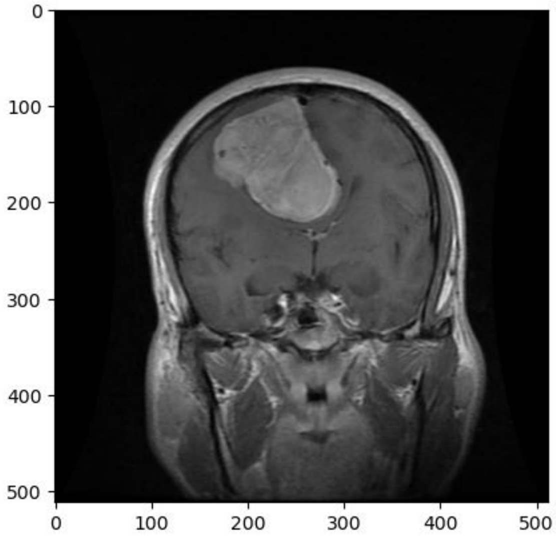
- Form a simplicial complex *K* from each image
  - Assign a vertex to each pixel, form 1-simplices if two pixels are adjacent, and form 2-simplices if three pixels are mutually adjacent
- Filtration on *K*
  - Let *V* be the set of vertices of *K* and $f : V \to \mathbb{R}$ be the pixel intensity.
  - Let f_{min}, f_{max} denote the minimum and maximum values of f.
  - Let $i_1 < i_2 < \cdots <$ be an increasing sequence of positive real numbers.
  - Let $K_i = \{S \in K : \forall v \in S, f(v) \leq i\}$. Then the sequence of subcomplexes $K_{fmin} \subseteq K_{fmin+i_1} \subseteq K_{fmin+i_2} \cdots \subseteq K_{fmax} = K$ is a filtration of *K*

K

MRI Scans

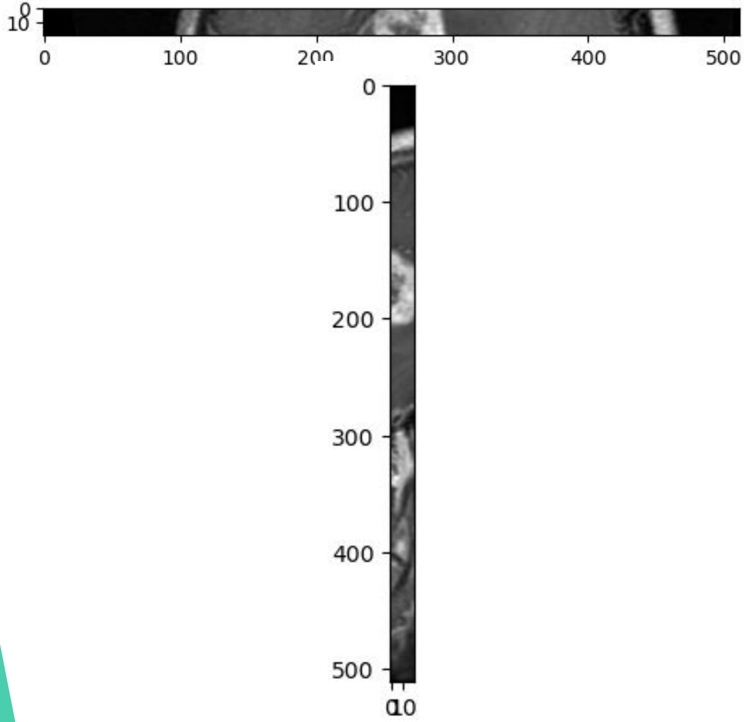


Glioma Tumor

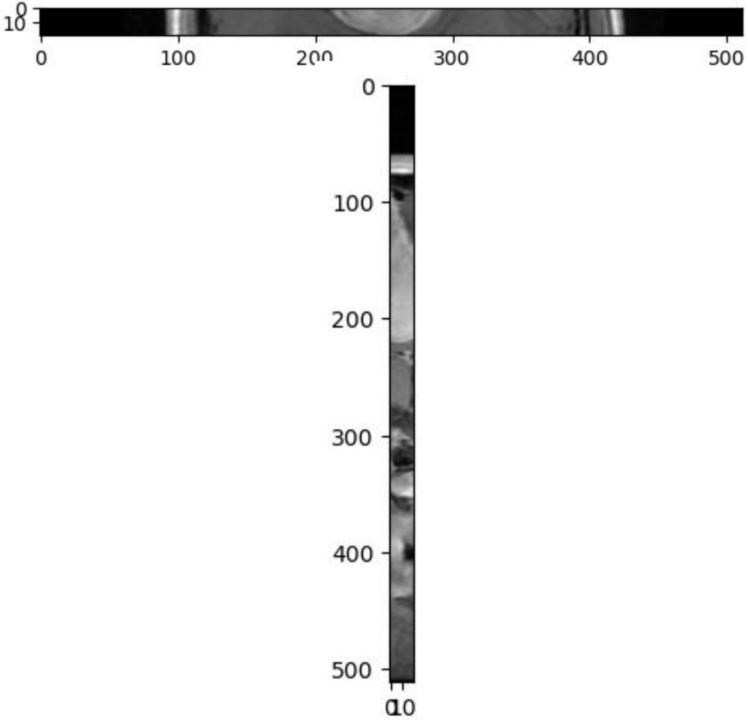


Meningioma Tumor

Slices

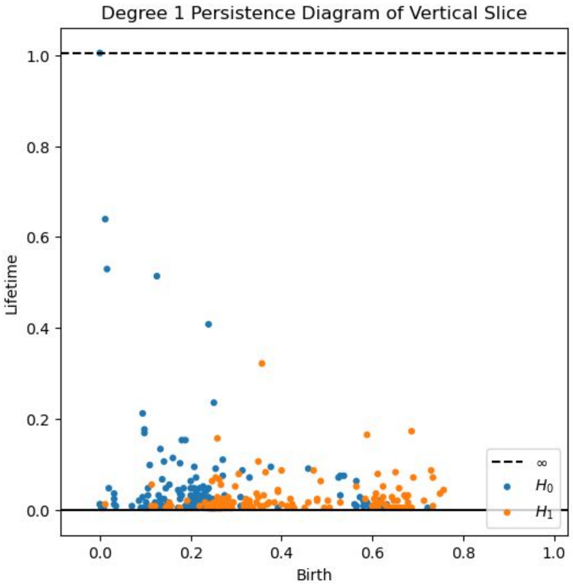


Glioma horizontal and vertical slice

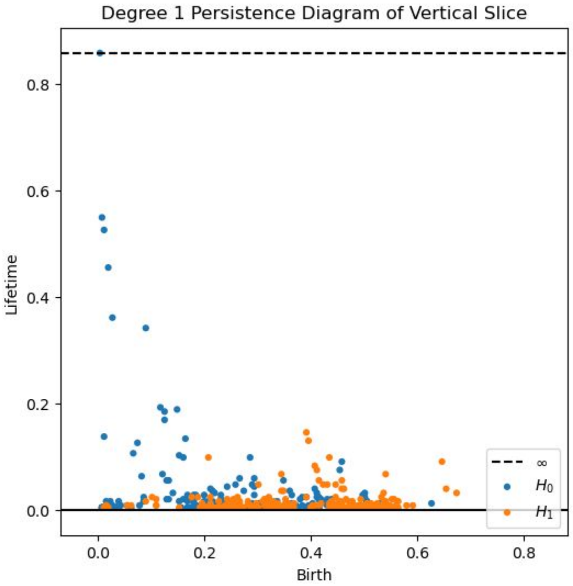


Meningioma horizontal and vertical slice

Persistence Diagrams

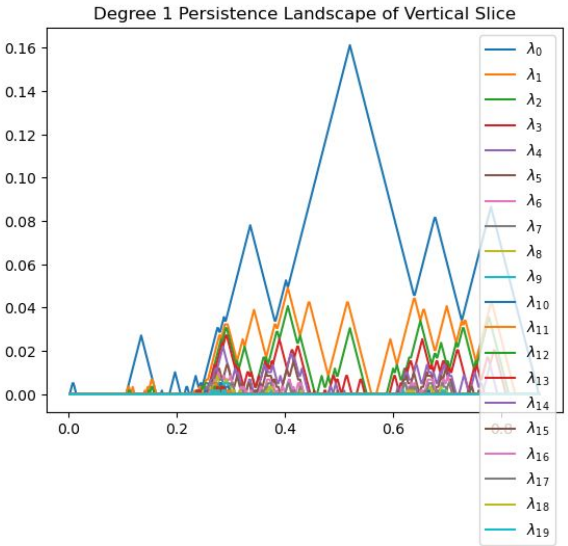


Glioma Persistence Diagram of a vertical slice

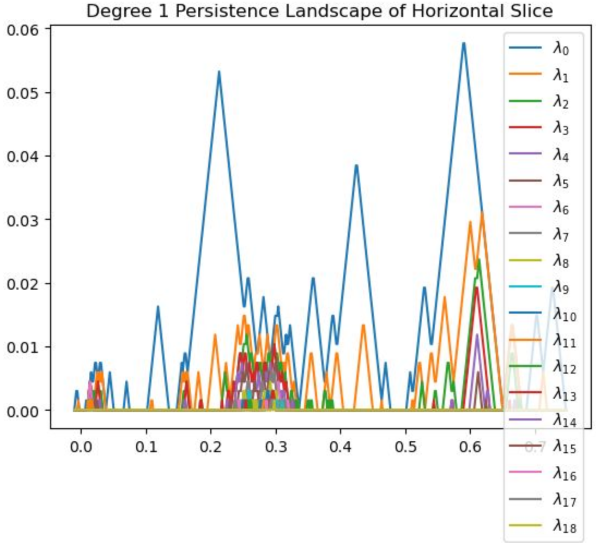


Meningioma Persistence Landscape of a vertical slice

Persistence Landscapes

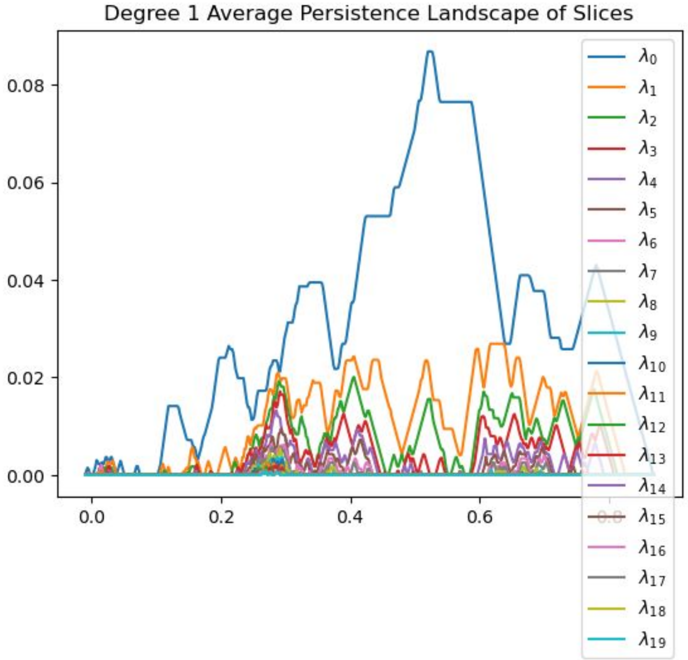


Glioma Persistence Landscape of a vertical slice

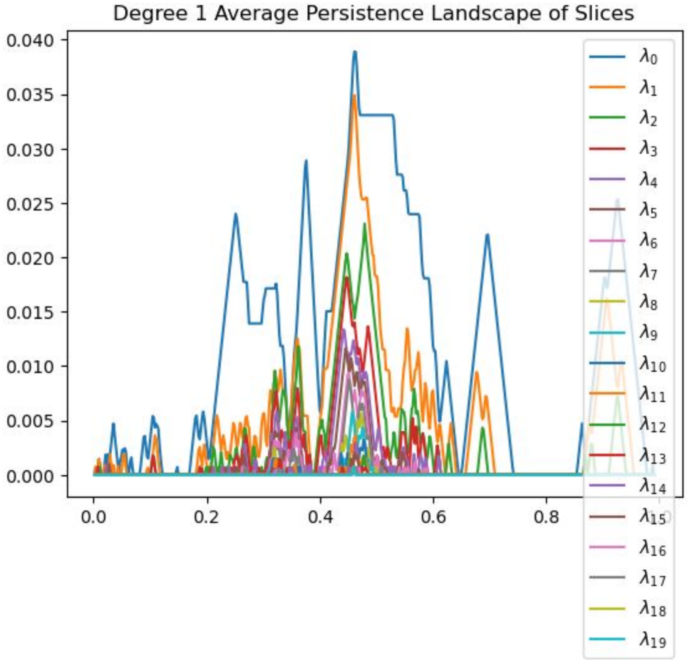


Glioma Persistence Landscape of a horizontal slice

Mean Persistence Landscape



Glioma Mean Persistence Landscape



Meningioma Mean Persistence Landscape

DNN Model Structure

- Used MLPClassifier from sklearn
- Model has 11 hidden layers
    - 5 layers of size 256, 4 layers of size 128, and 2 layers of size 64
- Constant learning rate = 0.001
- ReLU activation Function
- Adam Optimizer
- Used 600 training images (150 images from each tumor type), and 120 testing images (30 images from each tumor type)
    - Used limited training and testing data due to hardware constraints

```
MLPClassifier
MLPClassifier(hidden_layer_sizes=(256, 256, 256, 256, 256, 128, 128, 128, 128,
                                  64, 64),
              random_state=1)
```

Model Results

- Best Classifier Score of roughly 76%

```
clf = MLPClassifier(hidden_layer_sizes=(256,256,256,256,256,128,128,128,128,64,64), learning_rate = 'adaptive', random_state=1)
y_test = test_data_labels
y_pred=clf.predict(X_testscaled)
print(clf.score(X_testscaled, test_data_labels))
```

```
0.7583333333333333
```

Conclusion & Reflection

- What we achieved:
  - Best optimizer found to be Adam, best activation function found to be ReLU.
    - Training accuracy of 76%
- What we can improve on:
  - We have yet to find a suitable DNN to train our features on.
  - Best accuracy found by trial and error, modifying the hidden layer sizes, learning rate type, optimizer, and activation function.
  - We also hope to be able to train with more data for an overall better accuracy.

Thanks!