Goal for training:

- Use cognitively normal individuals, assuming brain is aging at roughly the same rate as their chronological ages
- Most relevant information for project:
    - Age at MRI
    - Clinical diagnosis

Note – data processing:

- Oasis data set has age at study entry; we need age at MRI
    - Image code includes days from study entry – must calculate age at MRI with this information
- Brain regions (MUSE ROIs) to include in graph nodes:
    - Cortical and subcortical gray matter tissue regions

Data for implementation of models:

- N by N adjacency matrix
    - N is the number of nodes/vertices in the graph
    - The adjacency matrix represents node-to-node connectivity of the graph
    - 'graph' is the space/domain where all data is defined
    - Common approach to define graph domain for brain:
        - Use intrinsic connectivity of the brain networks
            - Structural or functional (structural/functional connectomes)
        - Structural connectome matrix (like adjacency matrix)
            - Computed from diffusion weighted MRIs
            - In this case:
                - N is the number of anatomical brain regions (ROIs)
        - In graph-CNN implementations, we often assume a fixed graph for all subjects
            - i.e., fixing the domain where the data is defined, like classical CNN
            - attached paper: proposed atrophy data driven connectivity matrix estimation approach
- N by D feature matrix
    - D is number of features per node
    - Feature of interest is gray matter tissue volume of each ROI
- N by E label matrix
    - E is the number of classes
    - In our case, the outcome label is the chronological age

Validation plan:

- Waiting on epigenetic clock data – only available for ADNI cohort
- Use for independent validation
    - Train model with chron age as the outcome to learn, and only use cognitive healthy individuals in training
    - Then validate on their epigenetic clock estimates

- Another approach:
    - Train model directly on epigenetic clock estimates assuming these represent biological age
        - Might be a better estimate of brain age even in cognitively healthy individuals

Strossi columns - strossi_12_12_2019_14_14_58_oasis_apoestatus_adrcclinicaldataset

- ADRC_ADRCCLINICALDATA ID
- Subject
- Date
- Age
- mmse
- ageAtEntry
- cdr
- commun
- dx1, dx2, dx3, dx4, dx5
- homehobb
- judgment
- memory
- orient
- perscare
- apoe
- sumbox
- acsparnt
- height
- weight
- primStudy
- acsStudy

AIBL_ADOPIC_Demographics columns

- AIBL ID
- Visit
- Sex
- ApoE
- Diagnosis
- Age
- Amyloid Status
- Cohort

ADNI_muse results with icv columns (3 tables)

- Project
- Code
- Roi_idx
- Roi_volume

- Roi_name
- Roi_index

OASIS_muse results with icv columns (1 table)

- Project
- Code
- Roi_idx
- Roi_volume
- Roi_name
- Roi_index

AIBL_muse results with icv columns (1 table)

- Project
- Code
- Roi_idx
- Roi_volume
- Roi_name
- Roi_index

All the data has undergone – similar processing data pipelines, though from different

Looking at the adni tables, roi_name column, remove:

- Uppercase regions
- Corpus collosum - 161
- Cerebellum white matter -124-127
- 124-127
    - 124: Right Cerebellum Exterior
    - 125: Left Cerebellum Exterior
    - 126: Right Cerebellum White Matter
    - 127: Left Cerebellum White Matter
- 147-154
    - 147: frontal lobe WM right
    - 148: frontal lobe WM left
    - 149: occipital lobe WM right
    - 150: occipital lobe WM left
    - 151: parietal lobe WM right
    - 152: parietal lobe WM left
    - 153: temporal lobe WM right
    - 154: temporal lobe WM left

- 155-156
    - 155: fornix right
    - 156: fornix left
- 161
    - corpus callosum

- Remove icv however,
    - Take each individual and divide by icv to obtain '% of icv' for volumes
        - Want % volume of whole skull region / icv

Keep:

- 142-144
    - 142: Cerebellar Vermal Lobules I-V
    - 143: Cerebellar Vermal Lobules VI-VII
    - 144: Cerebellar Vermal Lobules VIII-X
- Lowercase
- Sentences
- Ventricles

ADNI demographics data:

PTDEGMOG:

- RID
- PTDOBMM
    - Ex: 4, 12, 1
    - M or MM
- PTDOBYY
    - YYYY
    - Ex: 1929, 1944

DXSUM_PDXCONV

- RID
- EXAMDATE
    - Ex: 9/29/2005
    - MM/DD/YYYY