Alzheimer’s Disease is the leading cause of dementia worldwide, and is characterized by the pathological deposition of tau and amyloid-beta protein aggregates in the cerebral cortex. I work in an Alzheimer’s Disease research lab studying the protein tau, so I’m particularly interested in the role of tau. In AD, tau misfolds into “neurofibrillary tangles” (NFTs), following a typical pattern of propagation known as the Braak stages. The Braak stage can only be formally diagnosed at autopsy, as it requires careful microscopic analysis of tissue from several brain regions by an experienced neuropathologist. However, the recent advent of a PET tracer specific to tau NFTs – known as 18F-AV1451 or Flortaucipir – enables *in cranio* visualization of regional tau accumulation over time.

The Alzheimer’s Disease Neuroimaging Initiative, a nationwide consortium that collects and analyzes neuroimaging and genetic data for thousands of individuals, recently published a dataset of longitudinal tau-PET neuroimaging data. In this dataset, tau-PET tracer uptake in over 100 brain regions was quantified for each scan using FreeSurfer, a powerful MRI segmentation program developed at the Martinos Center at Massachusetts General Hospital. In addition to PET scans, at each visit, patients were administered a series of neuropsychological tests to assess cognitive ability. While tau-PET has been shown in cross-sectional studies to associate with both postmortem Braak stage patterns and with cognitive status, longitudinal studies are sparse due to the newness of this specific tracer. Therefore, I would like to identify if the annual rate of tau accumulation in specific brain region(s) can predict changes in cognitive status.

In terms of how I will approach each stage:

Phase 1: I have identified a few papers examining the utility of tau-PET with machine learning to predict e.g. conversion from mild cognitive impairment (MCI) to dementia. I will read through each of these to understand what has been done already, what I can apply to my own project, and what novel findings I can contribute to the field.

Phase 2: I have downloaded the tau-PET and cognitive assessments from the ADNI website. The tau-PET dataset contains 1,121 entries, some of which contain multiple visits from one subject. The cognitive assessments include general cognitive diagnosis (either cognitively normal, mild cognitive impairment, or dementia) and a score for the Alzheimer's Disease Assessment—Cognitive 13-item scale (ADAS13), ranging from 0 to 70. These data are free to access, though one needs to register as a “Principal Investigator” on the website to gain access.

Phase 3: I intend to calculate the annual rate-of-change for regional tau-PET uptake as well as ADAS13 score. While general cognitive diagnosis cannot be quantified, I will summarise each subject as either stable (remaining at current cognitive status) or declining (i.e. cognitively normal to MCI, or MCI to dementia). I anticipate missing data in certain tau-PET regions of interest (ROIs) that are not detected in all subjects – for example, the ‘fifth ventricle” ROI is only reported in a small fraction of subjects. Depending on the overall missingness of a variable, I will either omit it or impute based on tau-PET values in the other ROIs.

Phase 4: I will explore two different models. First, I will investigate whether annual tau accumulation in different cortical regions can predict changes in the continuous ADAS-13 score variable. For this, I will use an ensemble model consisting of elastic net regression, kNN, and a neural network. I will also test dimensionality reduction with PCA as there are many cortical ROIs. Second, I will investigate whether regional tau accumulation can predict cognitive decline from one’s current cognitive status. As this is a binary classification problem (i.e. stable vs. declining), I will use logistic regression, SVM, and a neural network.

Phase 5: I will evaluate the continuous ADAS13 prediction model using MSE and correlation coefficient. I will evaluate the binary cognitive status prediction model using ROC and precision vs. recall. Additionally, I will use k-fold cross-validation using the caret package.

Phase 6: I will publish all code, analysis, and findings in a public GitHub repository. Additionally, I will create a Shiny app to interactively display the results of my analysis.