Development and validation of an interpretable deep learning framework for Alzheimer’s disease classification

Supplementary material

Neurologist approach to the ratings

Neurologist 1: When assessing MRI for structural changes concerning for AD, several key markers were assessed. These included hippocampal volume, entorhinal volume, temporal lobe atrophy, parietal lobe atrophy, ventricular atrophy, and whole brain atrophy. For each of these measures, the assessment will consider the degree of atrophy in each individual location compared to the brain as a whole. Part of the normal aging process includes diffuse atrophy that is not typically associated with cognitive impairments, so an assessment of atrophy of each brain region was assessed as to whether it was out of proportion to their age. In addition, the degree of atrophy in a specific region was compared to the degree of atrophy globally. Increased atrophy specifically in the medial temporal lobes or focal hippocampal atrophy which was out of proportion to the degree of global atrophy in the brain was noted as a finding concerning for AD pathology.

Neurologist 2: The MRI scans in all three planes (axial, coronal and sagittal) were reviewed to approximate degree of atrophy between sulci and atrophy of the hippocampus. White matter disease was not considered as a major factor in determining the symptoms or analysis since it was prominent only in a couple of scans.

Neurologist 3: Participant age and gender were reviewed first, followed by an initial review of the axial, coronal and sagittal T1-weighted MR images. The MMSE score was then viewed, followed by a second review of the MR images with maximum available clinical context before a yes/no prediction of AD dementia was made. Focal patterns of atrophy along with occasionally visualized evidence of possible normal pressure hydrocephalus, chronic small vessel disease or microhemorrhage were considered where applicable when reconciling an abnormal MMSE score with the possible presence of AD pathology. If a subject’s abnormal MMSE score and imaging were felt to be better explained by a condition other than AD (such as frontotemporal dementia, small vessel disease or normal pressure hydrocephalus), a designation of "No AD" was chosen.
Neurologist 4:
The goal was to diagnose AD cases from those who are cognitively normal (NC) from the following set of information: age, gender, MMSE result and T1-weighted MRI scans. Since the expectation is to distinguish between AD and NC cases, other forms of dementia such as vascular dementia or frontotemporal dementia were not considered. Only age, MMSE result and MRI scan information were considered. If the subjects had abnormal MMSE (cut-off of 26), the MRI scans were examined to see whether there was any hippocampal or cortical atrophy along with consideration of age. The following rules were generally used: (a) abnormal MMSE, atrophy, plus younger age, should be AD; (b) abnormal MMSE, atrophy, plus older age (>80 years), not explained by aging, should be AD; (c) borderline MMSE, 24-26, atrophy, plus younger age, should be AD; (d) for borderline MMSE, 24-26, plus older age, severity of the scans was evaluated more critically; (e) normal MMSE, plus younger age, and NO atrophy as NC cases; (f) normal MMSE, plus older age, some atrophy not noted as NC cases; (g) cases with normal MMSE and noted atrophy were carefully evaluated; for example a case with MMSE=27, 80yrs, and noted atrophy was given an AD label, and if MMSE=30 for the same scenario, then NC label was assigned.

Neurologist 5:
The MMSE score was considered along with age. An MMSE of 27-28 was considered normal for someone in their 80s, but not in their 50s or early sixties. Subsequently, the degree of atrophy (generalized and focally in the temporal lobes) on the MRI scans was reviewed. If mild generalized atrophy was observed in cases over 75 years of age, then they were considered as normal. However, if there was disproportionate atrophy of the temporal lobes, especially if MMSE was not normal, then the cases were rated as AD.

Neurologist 6:
The participant’s age and MMSE score were reviewed, followed by analysis of MRI scans for each case. MRIs were evaluated for any atrophy, especially within the hippocampal, medial temporal, as well as parietal and frontal lobes. Evidence of any enlarged ventricles, especially the enlargement of temporal horns, third and lateral ventricles was assessed. Correlation of the participant’s age with the degree of atrophy was then estimated. Additionally, other pathology that could explain a decline in MMSE such as evidence of white matter ischemic changes, prior strokes, encephalomalacia, was noted.

Neurologist 7:
Within the constraints of classifying AD based on the 4 variables available (age, gender, MMSE and MRI), MMSE was first considered for screening. Cases with normal or MCI (MMSE scores less than 25) were excluded. For those with low MMSE scores, the MRI scans were reviewed to make sure there was no evidence of obvious diseases. Since all available MRI sequences were T1-weighted, tumors, large strokes, and blood vessel abnormalities were reviewed. Age was considered as a factor, especially when the individual was very young to rule out the possibility of AD.

**Neurologist 8:**
MMSE score was first considered and cases with high scores (29 or 30) were evaluated as the ones with normal cognition, unless the MRI was considered quite abnormal. The MMSE was given the benefit of the doubt. If MMSE score was low (<26), then the cases were considered as abnormal. If the scores were intermediate (26-28), the MRI scans were more carefully evaluated. Age was used as a factor to assess the degree of atrophy. Gender was rarely considered.

**Neurologist 9:**
Age was first considered as a factor to screen the case. Subsequently, the MRI scan was reviewed to evaluate the global brain volume to see if it seemed appropriate for the individual’s age. Also, the hippocampal and medial temporal lobe areas were reviewed to see if they seemed appropriate for overall brain volume or if they seemed atrophic and out of proportion to the rest of the brain.

**Neurologist 10:**
The MRI scans for each case were reviewed to see whether there were atrophies of the cortical regions, especially atrophy in the parietal lobe. If the degree of the atrophy did not match the age of the case, and if there were hippocampus atrophies at the same time, then these features may indicate AD. These observations in conjunction with the MMSE score were considered to make the assessment. If there were cerebral small vessel diseases or frontotemporal lobe atrophies, the diagnosis may need careful consideration.

**Neurologist 11:**
MRI scans were reviewed first and an initial diagnosis for AD (including preclinical, MCI, dementia) was made. After reviewing the MRIs, age and MMSE were considered and AD ratings were adjusted, if needed. On the MRIs, evidence of confounding diagnosis such as ischemic strokes or tumors were evaluated. Evidence to support AD such as medial temporal lobe atrophy, and entorhinal cortex atrophy were considered. Moreover, the medial temporal lobe atrophy
scale (MTAS), was used. Higher MTAS indicated more atrophy and higher likelihood of AD. Other imaging features that were noted were generalized cerebral atrophy and white matter lesions. White matter lesions are quite common in elderly participants; however, it may represent small vessel disease in some cases. The STRIVE criteria and Fazekas score for small vessel disease were also used. In participants with low MMSE scores with intact medial temporal lobe structures, small vessel disease can explain decreased cognition without a diagnosis of AD. For example, MTAS grade 1-2 were not considered as AD for participants older than 75 years.

Summary of neurologist approaches

Participating neurologists were requested to explain their reasoning when tasked to predict AD status from collections of MRI scans, age, gender, and MMSE score for a subset of participants with masked disease diagnosis. While there was notable variation in the order in which imaging and non-imaging data were considered, the two forms of information were widely considered complementary. Focal atrophy of key cerebral regions (notably the hippocampus and temporal lobes) was considered in light of generalized age-appropriate atrophy, and imaging was broadly utilized to rule out competing etiologies of dementia such as frontotemporal degeneration and vascular disease. MMSE, often considered in the context of age, was widely employed as comparison to salient imaging features as well. Collectively, these perspectives speak to the importance of an integrated approach to AD diagnosis in which distinct forms of information are reconciled prior to an ultimate classification of disease status.