

# **Supplementary material**

**AI-based differential diagnosis of dementia etiologies on multimodal data**

Dataset (group)	MMSE mean $\pm$ std	MOCA mean $\pm$ std	APOE N, % positive
NACC			
NC [n = 17242]	28.76 $\pm$ 1.57 <sup>^</sup>	26.28 $\pm$ 2.7 <sup>^</sup>	3852, 30.07% <sup>^</sup>
MCI [n = 7582]	26.94 $\pm$ 2.59 <sup>^</sup>	22.61 $\pm$ 3.53 <sup>^</sup>	1913, 38.2% <sup>^</sup>
AD [n = 16131]	18.94 $\pm$ 6.79 <sup>^</sup>	15.15 $\pm$ 5.78 <sup>^</sup>	6840, 56.05% <sup>^</sup>
LBD [n = 1913]	19.7 $\pm$ 7.07 <sup>^</sup>	16.13 $\pm$ 5.95 <sup>^</sup>	600, 44.02% <sup>^</sup>
VD [n = 1919]	19.56 $\pm$ 6.85 <sup>^</sup>	16.19 $\pm$ 5.6 <sup>^</sup>	577, 42.06% <sup>^</sup>
PRD [n = 114]	13.37 $\pm$ 10.03 <sup>^</sup>	16.18 $\pm$ 7.6 <sup>^</sup>	6, 23.08% <sup>^</sup>
FTD [n = 2898]	19.09 $\pm$ 8.13 <sup>^</sup>	17.06 $\pm$ 6.66 <sup>^</sup>	740, 31.97% <sup>^</sup>
NPH [n = 138]	20.54 $\pm$ 6.46 <sup>^</sup>	17.38 $\pm$ 5.1 <sup>^</sup>	38, 42.7% <sup>^</sup>
SEF [n = 808]	20.63 $\pm$ 6.33 <sup>^</sup>	17.21 $\pm$ 5.46 <sup>^</sup>	225, 39.4% <sup>^</sup>
PSY [n = 2700]	20.33 $\pm$ 6.52 <sup>^</sup>	16.15 $\pm$ 5.83 <sup>^</sup>	911, 46.13% <sup>^</sup>
TBI [n = 265]	21.07 $\pm$ 6.29 <sup>^</sup>	17.24 $\pm$ 6.82 <sup>^</sup>	74, 39.15% <sup>^</sup>
ODE [n = 1234]	20.32 $\pm$ 6.99 <sup>^</sup>	17.11 $\pm$ 5.89 <sup>^</sup>	403, 42.38% <sup>^</sup>
<i>p-value</i>	<1.0e-200	<1.0e-200	<1.0e-200
NIFD			
NC [n = 124]	29.35 $\pm$ 0.76	27.58 $\pm$ 1.53 <sup>^</sup>	N.A.
FTD [n = 129]	24.75 $\pm$ 4.54 <sup>^</sup>	19.69 $\pm$ 5.72 <sup>^</sup>	N.A.
<i>p-value</i>	1.961e-23	2.645e-16	N.A.
PPMI			
NC [n = 171]	N.A.	27.51 $\pm$ 2.37 <sup>^</sup>	N.A.
MCI [n = 27]	N.A.	24.69 $\pm$ 3.27 <sup>^</sup>	N.A.
<i>p-value</i>	N.A.	3.004e-07	N.A.
AIBL			
NC [n = 480]	28.7 $\pm$ 1.24	N.A.	12, 2.5%
MCI [n = 102]	27.1 $\pm$ 2.08	N.A.	12, 11.76%
AD [n = 79]	20.42 $\pm$ 5.46	N.A.	14, 17.72%
<i>p-value</i>	4.585e-121	N.A.	8.951e-09
OASIS			
NC [n = 424]	28.99 $\pm$ 1.25 <sup>^</sup>	N.A.	140, 33.02%
MCI [n = 27]	28.15 $\pm$ 1.67	N.A.	11, 40.74%
AD [n = 32]	23.91 $\pm$ 5.05	N.A.	20, 62.5%
LBD [n = 4]	25.5 $\pm$ 2.69	N.A.	2, 50.0%
FTD [n = 4]	18.33 $\pm$ 8.26 <sup>^</sup>	N.A.	4, 100.0%
<i>p-value</i>	4.439e-50	N.A.	4.510e-05
LBDSU			
NC [n = 134]	N.A.	27.43 $\pm$ 2.23 <sup>^</sup>	N.A.
MCI [n = 35]	N.A.	24.0 $\pm$ 3.14	N.A.
LBD [n = 13]	N.A.	16.69 $\pm$ 4.75	N.A.
<i>p-value</i>	N.A.	2.231e-30	N.A.
4RTNI			
NC [n = 12]	27.2 $\pm$ 2.4 <sup>^</sup>	24.2 $\pm$ 2.44 <sup>^</sup>	N.A.
MCI [n = 31]	26.1 $\pm$ 3.95 <sup>^</sup>	21.19 $\pm$ 4.83 <sup>^</sup>	N.A.
FTD [n = 37]	21.49 $\pm$ 7.2 <sup>^</sup>	17.14 $\pm$ 7.4 <sup>^</sup>	N.A.
<i>p-value</i>	1.657e-03	3.605e-03	N.A.
ADNI			
NC [n = 481]	29.05 $\pm$ 1.12	25.71 $\pm$ 2.59 <sup>^</sup>	138, 29.61% <sup>^</sup>
MCI [n = 971]	27.62 $\pm$ 1.81	23.18 $\pm$ 3.23 <sup>^</sup>	438, 47.2% <sup>^</sup>
AD [n = 369]	23.19 $\pm$ 2.11	16.8 $\pm$ 4.5 <sup>^</sup>	229, 64.33% <sup>^</sup>
<i>p-value</i>	<1.0e-200	1.010e-116	3.117e-22
FHS			
NC [n = 394]	26.05 $\pm$ 3.36 <sup>^</sup>	N.A.	N.A.
MCI [n = 434]	25.2 $\pm$ 3.4 <sup>^</sup>	N.A.	N.A.
AD [n = 687]	21.63 $\pm$ 5.08 <sup>^</sup>	N.A.	N.A.
LBD [n = 73]	22.91 $\pm$ 4.16 <sup>^</sup>	N.A.	N.A.
VD [n = 113]	22.41 $\pm$ 5.44 <sup>^</sup>	N.A.	N.A.
FTD [n = 8]	20.33 $\pm$ 3.4 <sup>^</sup>	N.A.	N.A.
<i>p-value</i>	3.132e-26	N.A.	N.A.

Table S1: **Study population.** Nine independent datasets were used for this study, including ADNI, NACC, NIFD, PPMI, OASIS, LBDSU, 4RTNI, and FHS. Data from NACC, NIFD, PPMI, OASIS, LBDSU, and 4RTNI were used for model training. Data from ADNI, FHS, and a held-out set from NACC were used for model testing. The *p*-value for each dataset indicates the statistical significance of inter-group differences per column. We used one-way ANOVA and  $\chi^2$  tests for continuous and categorical variables, respectively. Please refer to Glossary 1 for more information on the acronyms. Here N.A. denotes not available. The symbol <sup>^</sup> indicates that data was not available for some subjects.

Cohort	Features
ADNI	<ul style="list-style-type: none"> <li>• Subject's age at visit</li> <li>• Subject' sex</li> <li>• Years of education</li> <li>• Hispanic/Latino ethnicity</li> <li>• Derived NIH race definitions</li> <li>• Total years smoked cigarettes</li> <li>• Heart attack/cardiac arrest</li> <li>• Hypertension</li> <li>• Stroke</li> <li>• Depression episodes more than two years ago</li> <li>• Other psychiatric disorder</li> <li>• Indicator of first-degree family member with cognitive impairment</li> <li>• Alcohol abuse - clinically significant occurring over a 12-month period manifested in one of the following areas: work, driving, legal, or social</li> <li>• Agitation or aggression severity</li> <li>• Motor disturbance severity</li> <li>• Delusions severity</li> <li>• Disinhibition severity</li> <li>• Hallucinations severity</li> <li>• Depression or dysphoria severity</li> <li>• Nighttime behavior severity</li> <li>• Apathy or indifference severity</li> <li>• Elation or euphoria severity</li> <li>• Anxiety severity</li> <li>• Appetite and eating severity</li> <li>• Irritability or lability severity</li> <li>• Number of APOE e4 alleles</li> <li>• Digit span forward trials correct</li> <li>• Digit span backward trials correct</li> <li>• Digit span forward length</li> <li>• Digit span backward length</li> <li>• Total MMSE score (using D-L-R-O-W)</li> <li>• Trail making test Part A - Total number of seconds to complete</li> <li>• Trail making test Part B - Total number of seconds to complete</li> <li>• Logical memory IIA - Delayed - Total number of story units recalled</li> </ul> <ul style="list-style-type: none"> <li>• Total number of story units recalled from this current test administration</li> <li>• Animals - Total number of animals named in 60 seconds</li> <li>• MoCA total score - corrected for education</li> <li>• Boston naming test (30) - Total score</li> <li>• Total GDS score</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Assembling tax records, business affairs, or other paper</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Shopping alone for clothes, household necessities, or groceries</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Writing checks, paying bills, or balancing a checkbook</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Heating water, making a cup of coffee, turning off the stove</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Playing a game of skill such as bridge or chess, working on a hobby</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Traveling out of the neighborhood, driving, or arranging to take public transportation</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Paying attention to and understanding a TV program, book, or magazine</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Preparing a balanced meal</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Keeping track of current events</li> <li>• In the past four weeks, did the subject have any difficulty or need help with: Remembering appointments, family occasions, holidays, medications</li> <li>• Imaging (MRI scans)</li> </ul>
FHS	<ul style="list-style-type: none"> <li>• Subject's sex</li> <li>• Subject's age at visit</li> <li>• Hispanic/Latino ethnicity</li> <li>• Race</li> <li>• Derived NIH race definitions</li> <li>• Marital status</li> <li>• Left- or right-handedness</li> <li>• Subject's weight (lbs)</li> <li>• Subject's height (inches)</li> <li>• Body mass index (BMI)</li> <li>• Blood pressure (sitting), systolic</li> <li>• Blood pressure (sitting), diastolic</li> <li>• Smoked cigarettes in last 30 days</li> <li>• Total MMSE score (using D-L-R-O-W)</li> <li>• Boston naming test (30) - Total score</li> <li>• History of stroke</li> <li>• Reported current use of a diabetes medication</li> <li>• Reported current use of lipid lowering medication</li> <li>• Imaging (MRI scans)</li> </ul>

Table S2: **Features from the ADNI and FHS cohorts.** The table lists all the features that were obtained from the ADNI and FHS cohorts. Note that ADNI and FHS were used as external datasets for model testing.

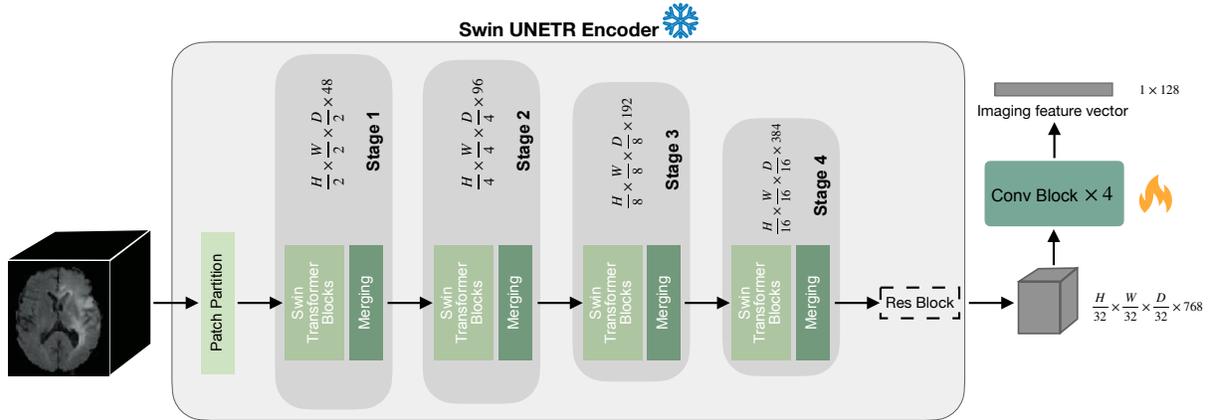
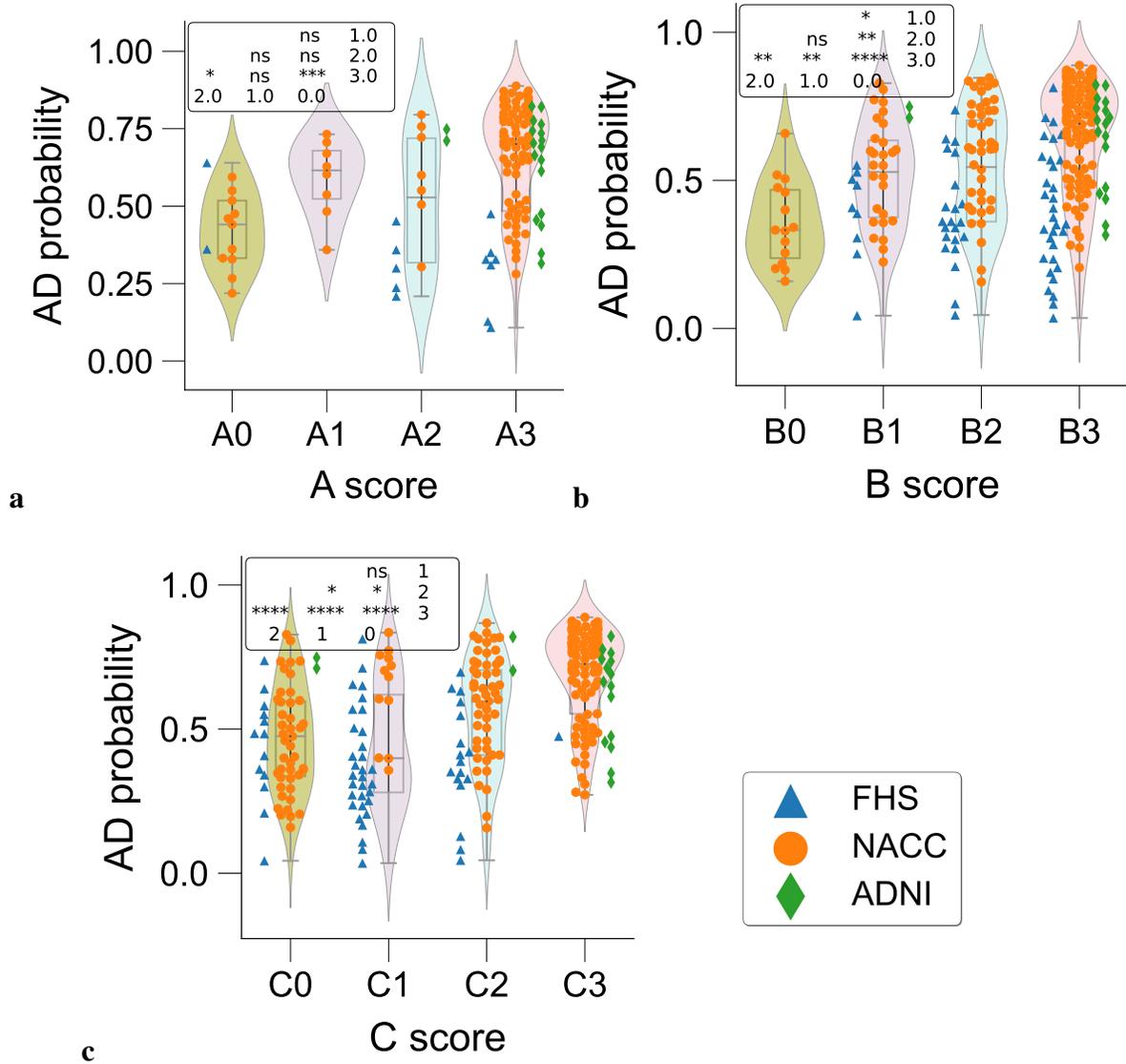


Figure S1: **Image feature extraction.** The Swin UNETR encoder, utilizing pre-trained weights, was leveraged to extract image embeddings from multi-sequence MRI scans into a latent space representation. Subsequently, these embeddings underwent a series of downsampling convolutional operations to achieve a condensed token dimension of  $1 \times 128$ . This dimensional reduction facilitated a consistent input format for both imaging and non-imaging data into the backbone transformer. Within this architecture, the Swin UNETR encoder’s weights remained static (frozen), ensuring the integrity of the pre-trained features, while the downsampling blocks were subject to optimization during the training phase, allowing for adaptive learning of the imaging feature vector.

Dataset (group)	T1	T2	FLAIR	OTHER
NACC	2313	365	465	629
NIFD	633	414	537	1811
PPMI	241	N.A.	N.A.	N.A.
AIBL	681	N.A.	334	N.A.
OASIS	665	N.A.	N.A.	N.A.
LBDSU	181	N.A.	N.A.	N.A.
4RTNI	165	119	120	N.A.
ADNI	1189	N.A.	N.A.	N.A.
FHS	135	126	128	187

Table S3: **MRI sequences used for model development.** T1-weighted, T2-weighted, fluid attenuated inversion recovery and other sequences including susceptibility weighted imaging, and diffusion-weighted imaging, among others, were included from NACC, NIFD, PPMI, AIBL, OASIS, LBDSU and 4RTNI for model training. A portion of MRIs from the NACC dataset along with MRIs from ADNI, and FHS were reserved for model testing.



**Figure S2: Neuropathological validation.** Array of violin plots with integrated box plots, delineating the probability distributions as predicted by the model for various stages of the Thal phase for amyloid plaques (A score), Braak stage for neurofibrillary degeneration (B score), and the density of neocortical neuritic plaques (CERAD or C score). The analysis encompasses data from three distinct cohorts: the Framingham Heart Study (FHS), the National Alzheimer’s Coordinating Center (NACC), and the Alzheimer’s Disease Neuroimaging Initiative (ADNI), each denoted by unique markers (triangles, circles, and diamonds, respectively). Statistical significance is encoded using asterisks determined by Dunn-Bonferroni post-hoc test: one asterisk (\*) for a p-value less than 0.05; two asterisks (\*\*) for p-values less than 0.01; three asterisks (\*\*\*) for p-values less than 0.001; and four asterisks (\*\*\*\*) for p-values less than 0.0001, reflecting increasing levels of statistical significance.

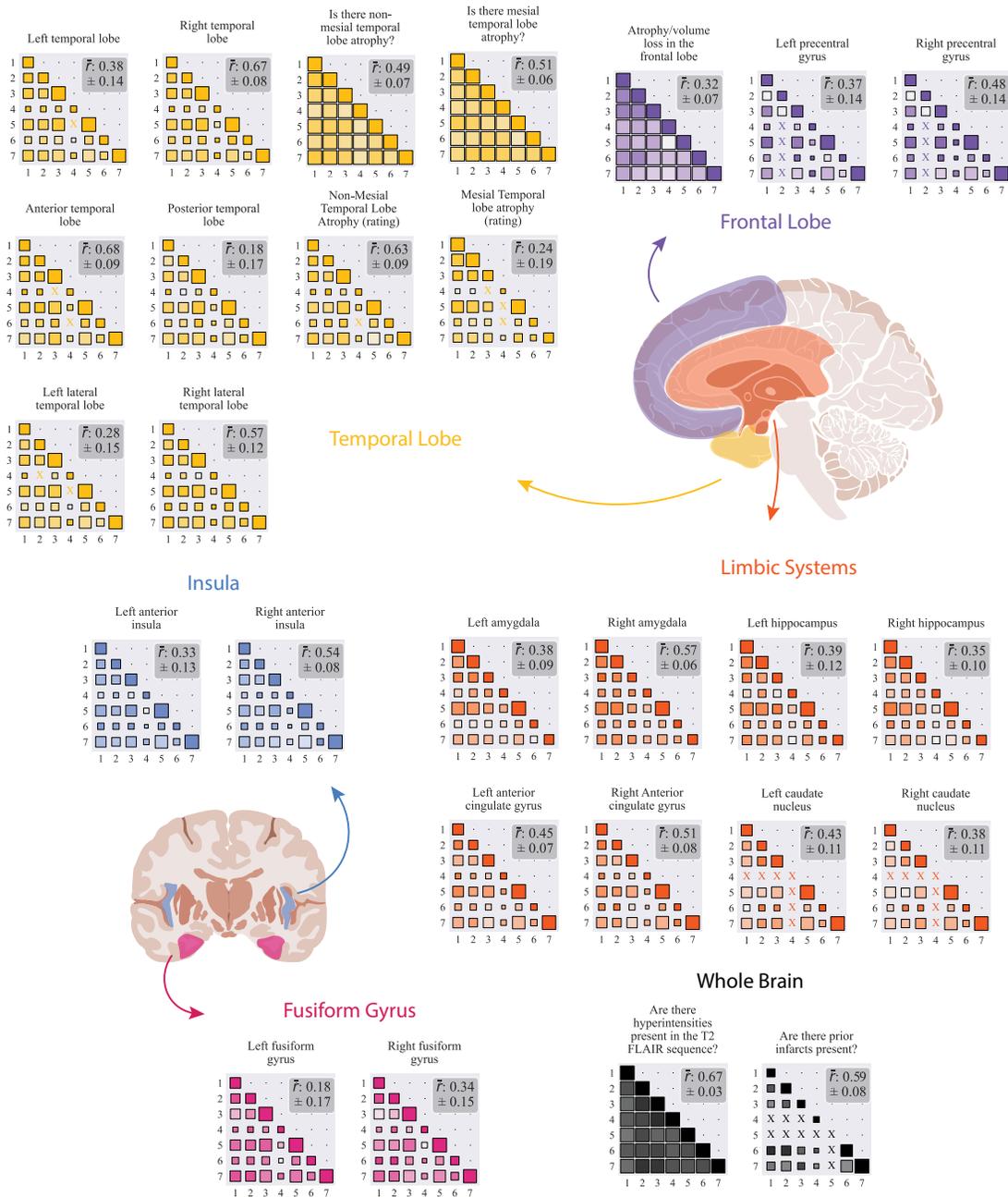


Figure S3: **Radiologist-driven assessments of neurodegeneration.** The graphic depicts a Pearson correlation analysis of independent evaluations by seven radiologists on a set of 70 cases, focusing on specific queries and excluding any case labeled as ‘Cannot Assess’ by any radiologist. In the matrix, the intensity of color within each cell indicates the strength of the Pearson correlation coefficient, while the size of the cell signifies the count of cases that were jointly evaluated by each pair of radiologists. An ‘X’ marks the pairs for which one or both provided assessments on a subset of valid cases that displayed no variance, thereby precluding the calculation of a Pearson correlation. Accompanying the matrix, the average pairwise Pearson correlation coefficient for each assessed region is displayed, complete with a 95% confidence interval.

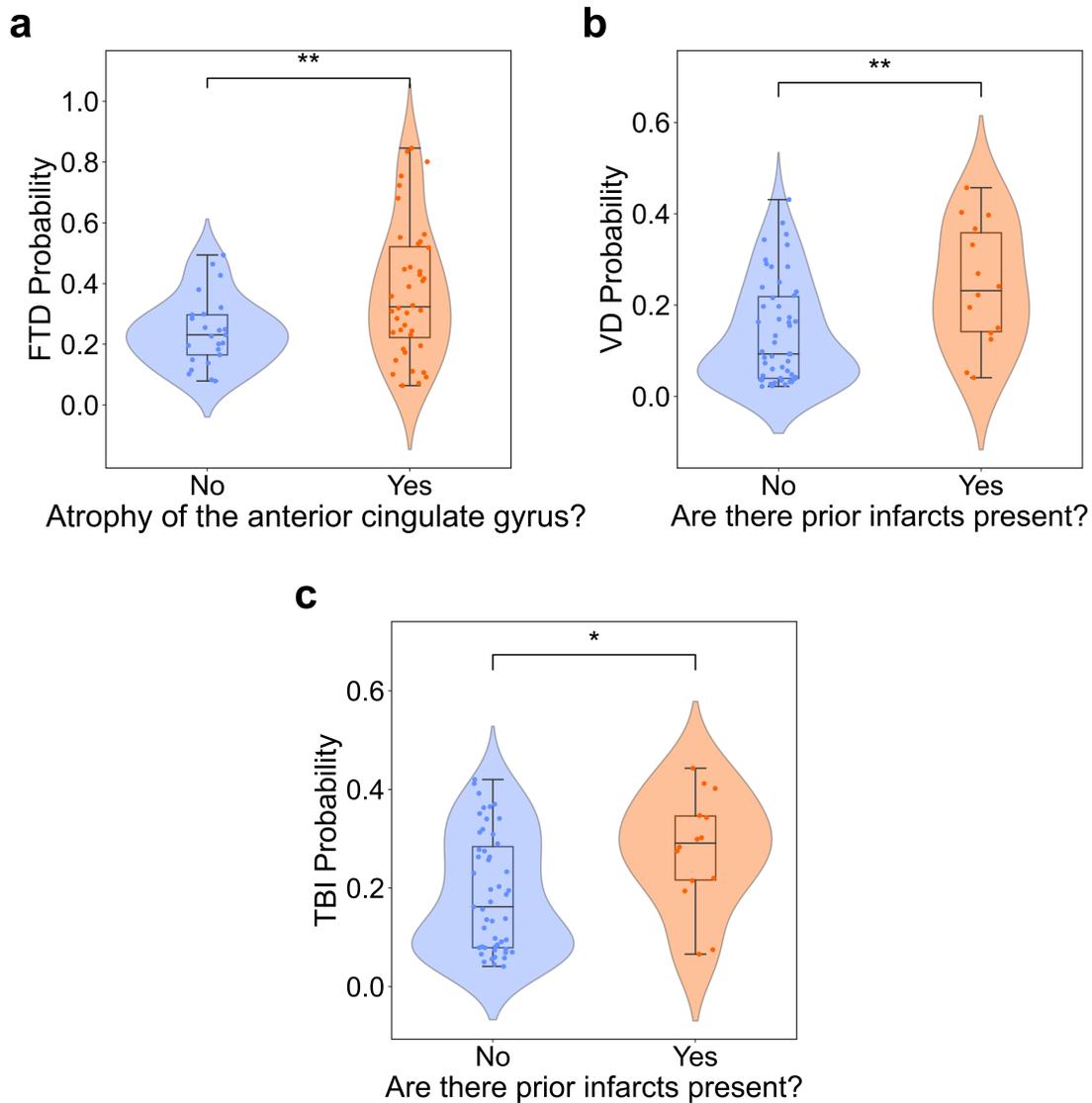


Figure S4: **Radiologist review of neurodegeneration markers.** The figure employs violin plots with embedded box plots to display the distribution of the model confidence scores for various etiologies, segmented by the consensus expert assessment for various key diagnostic questions that identify markers relevant to frontotemporal, vascular, and traumatic brain injury-induced dementia. Statistical significance is annotated within the figure: a single asterisk (\*) denotes  $p < 0.05$ ; and two asterisks (\*\*) signify  $p < 0.01$ , with these levels of significance determined through the Dunn-Bonferroni post-hoc test.

Cohort	Age at death (years) mean $\pm$ std	Male gender (percentage)	CDR mean $\pm$ std
NACC			
AD [n = 131]	79.28 $\pm$ 10.61	74, 56.49%	1.44 $\pm$ 0.75
LBD [n = 44]	76.64 $\pm$ 8.31	36, 81.82%	1.44 $\pm$ 0.66
VD [n = 15]	83.73 $\pm$ 10.39	7, 46.67%	1.37 $\pm$ 0.86
PRD [n = 5]	562.6 $\pm$ 4.96	5, 100%	0.7 $\pm$ 0.24
FTD [n = 20]	67.15 $\pm$ 8.61	13, 65%	1.82 $\pm$ 0.98
NPH [n = 1]	77 $\pm$ 0.0	1, 100%	0.5 $\pm$ 0.0
SEF [n = 3]	73 $\pm$ 24.04	2, 66.67%	0.83 $\pm$ 0.23
TBI [n = 1]	87 $\pm$ 0.0	1, 100%	2.0 $\pm$ 0.0
ODE [n = 11]	68.82 $\pm$ 14.99	7, 63.64%	1.45 $\pm$ 0.86
<i>p-value</i>	1.428e-08	5.956e-02	1.471e-01
ADNI			
AD [n = 19]	82.05 $\pm$ 9.03	13, 68.42%	0.86 $\pm$ 0.22
FHS			
AD [n = 55]	91 $\pm$ 7.0	18, 32.72%	1.01 $\pm$ 0.15
LBD [n = 4]	92.75 $\pm$ 2.16	3, 75%	1.0 $\pm$ 0.0
VD [n = 5]	93 $\pm$ 3.35	4, 80%	1.2 $\pm$ 0.4
FTD [n = 2]	79 $\pm$ 4.0	1, 50%	1.0 $\pm$ 0.0
<i>p-value</i>	1.584e-01	8.227e-02	2.972e-01

Table S4: **Cases with post mortem findings used for model validation.** Model validation was conducted using cases with post-mortem findings from three independent datasets: ADNI, NACC, and FHS. Continuous variables were analyzed using one-way ANOVA, while categorical variables were assessed with  $\chi^2$  tests. The p-values derived for each dataset reflect the statistical significance of differences between groups for each column.

Dataset (group)	Balanced Accuracy	Precision	Sensitivity	Specificity	F1 Score	MCC	AUROC	AUPR
NACC								
NC	0.92	0.92	0.89	0.95	0.90	0.85	0.98	0.96
MCI	0.83	0.53	0.80	0.86	0.64	0.57	0.91	0.67
DE	0.94	0.92	0.94	0.93	0.93	0.87	0.99	0.98
AD	0.89	0.84	0.89	0.90	0.86	0.78	0.96	0.93
LBD	0.86	0.45	0.77	0.96	0.57	0.56	0.96	0.68
VD	0.83	0.30	0.74	0.92	0.42	0.43	0.94	0.48
PRD	0.86	0.027	0.76	0.95	0.052	0.14	0.96	0.074
FTD	0.89	0.38	0.87	0.91	0.53	0.54	0.96	0.65
NPH	0.70	0.018	0.48	0.92	0.035	0.081	0.88	0.02
SEF	0.66	0.065	0.44	0.88	0.11	0.13	0.82	0.063
PSY	0.79	0.26	0.69	0.88	0.38	0.37	0.89	0.36
TBI	0.71	0.025	0.56	0.87	0.048	0.096	0.85	0.033
ODE	0.71	0.084	0.62	0.81	0.15	0.17	0.83	0.10
ADNI								
NC	0.83	0.52	0.97	0.68	0.68	0.58	0.9	0.66
MCI	0.69	0.86	0.48	0.91	0.61	0.42	0.79	0.80
DE	0.88	0.60	0.90	0.85	0.72	0.66	0.95	0.86
AD	0.88	0.69	0.86	0.9	0.76	0.70	0.94	0.84
FHS								
NC	0.58	0.30	0.61	0.55	0.40	0.14	0.62	0.30
MCI	0.51	0.58	0.032	0.99	0.061	0.088	0.55	0.32
DE	0.61	0.68	0.42	0.80	0.52	0.24	0.68	0.68
AD	0.61	0.63	0.38	0.84	0.47	0.25	0.69	0.62
LBD	0.51	0.095	0.027	0.99	0.043	0.028	0.59	0.062
VD	0.66	0.24	0.42	0.90	0.31	0.26	0.71	0.31
FTD	0.52	0.007	0.12	0.92	0.014	0.01	0.51	0.008

Table S5: **Model performance.** This table displays a set of performance metrics produced by the model across the NACC, ADNI, and FHS datasets. Demographic information for each cohort can be found in Table S1.

## **Neurologist approach to the ratings**

**Neurologist 1** The clinical data was reviewed initially, taking note of potential contributors such as extreme age or education (e.g., age > 90, education less than 9 grades), primary language, and language of cognitive testing. Pertinent factors like a history of TIA or Stroke, Parkinson's disease (PD) diagnosis and/or PD medication usage, known genetic mutations, closed head injury, alcohol or substance use disorders, chronic psychiatric symptoms/disorders, and APOE genotype were assessed. Next, the current level of functional abilities was evaluated from the provided initial description (e.g., independent living, requiring assistance with some or all activities) and Functional Activities Questionnaire (FAQ) responses. FAQ scores of 9 or higher typically indicated significant limitations with instrumental activities of daily living, supporting a dementia diagnosis. FAQ scores ranging from 4 to 8 would align with mild cognitive impairment (MCI) if cognitive test scores indicated cognitive decline. Subsequently, cognitive test scores were reviewed, with focus on age, education, and gender-adjusted z-scores. For those with normal cognition, no z-scores deviated by 1 standard deviation below the mean (i.e., no score of -1.0 or worse). Persons with MCI would exhibit at least one z-score of -1.5 or worse (e.g., -1.75) or two scores of -1.0 in the same cognitive domain. Persons with dementia would typically present with two or more scores at -2.0 or worse. Interpretation for patients with very low education or non-native language cognitive testing was approached cautiously. Following this, brain MRIs (T1-weighted images) were reviewed for signs of atrophy, the pattern of atrophy, and cerebrovascular disease. When available, diffusion-weighted imaging was used to identify a diffusion restriction pattern commonly seen in prion diseases. Functional abilities and cognitive test scores were used to classify persons as normal, MCI, or dementia. For persons between categories, a continuum scale was employed. For instance, a score of 80 for MCI and 20 for dementia would indicate an 80% likelihood of classification as MCI and a 20% likelihood of classification as dementia. For individuals with MCI or dementia, the most likely diagnostic category or categories were selected. In cases of mixed dementia or unclear causation, multiple diagnostic categories were chosen, with their scores summing to 100. Each category's score reflected the estimated contribution and, for mixed dementias, the extent of their contribution. For example, a score of 70 for Alzheimer's disease (AD), 20 for Lewy body dementia (LBD), and 10 for vascular dementia would signify an estimated 70% contribution from AD, 20% from LBD, and 10% from cerebrovascular disease.

**Neurologist 2** The evaluation of case reports began with a comprehensive analysis of demographics, available medical history, apoE4 status, structured family history, and an assessment of the patient's level of functional independence. Subsequently, a thorough examination of corresponding clinical scales and neuropsychological test results was conducted. Careful observations were made regarding the subject's educational background, the presence of visual or hearing impairments, and whether the tests were conducted in the subject's native language. Following this, the synthesis of clinical data allowed for the prediction of the presence of mild cognitive impairment (MCI), dementia, or cognitive states falling below the MCI threshold, often referred to as 'normal' cognition. These predictions were quantified, with the most probable diagnosis assigned a rating exceeding 50%, while the others received lower ratings, reflecting the confidence in the diagnosis. Subsequently, the MRI sequences were examined alongside the case report to identify

factors contributing to the patient's clinical condition. Distinctly, findings such as medial temporal atrophy and parietal atrophy were prominently associated with Alzheimer's disease, whereas the presence of flair hyperintensity and focal encephalomalacia without an alternative cause was considered indicative of vascular burden and/or dementia, especially when accompanied by deep and/or brainstem microhemorrhages. Brainstem atrophy was frequently observed in cases suggestive of potential stroke or Lewy body conditions, and the use of DWI sequences allowed for the potential identification of conditions like prion disease and epilepsy-related disorders. In assessing the clinical significance of these contributors, the most plausible factors were rated highest, while other contributors received lower but still significant ratings, typically exceeding 50%. However, distinguishing psychiatric features stemming from a neurodegenerative process from those arising as independent comorbid issues occasionally posed a challenge. Importantly, observed vascular burden in imaging, even when it didn't independently warrant a dementia diagnosis, was consistently acknowledged under the vascular category, often rated highly due to the confidence in its clinical significance.

**Neurologist 3** In the approach to differential diagnosis for dementia, a detailed case overview encompassed a wide spectrum of clinical information including demographics, vitals, and comprehensive personal and medical histories, alongside results from systematic physical, neurological, psychiatric, and neuro-cognitive evaluations. Cognitive function was assessed using clinician impressions from neuropsychiatric evaluations and standardized testing with MMSE or MoCA, facilitating the distinction between normal cognition, mild cognitive impairment, and dementia. Functional assessments provided insights into the impact of neurological disorders on daily living activities. Specific scales and questionnaires, such as the Hachinski Ischemic Score, evaluations for Progressive Supranuclear Palsy (PSP), and Corticobasal Syndrome (CBS), the Unified Parkinson's Disease Rating Scale, and the Neuropsychiatric Inventory Questionnaire, were instrumental in identifying localized or generalized neurological deficits, signs and symptoms of Parkinson's disease and related conditions, and characteristic features of Lewy body dementia, such as visual hallucinations. The presence of typical symptoms for disorders like normal pressure hydrocephalus also contributed to fine-tuning the differential diagnosis. The Geriatric Depression Scale was employed to discern if primary psychiatric disorders might mimic dementia presentations. An extensive review of neurocognitive testing data aided in differentiating Alzheimer's disease from other cognitive disorders. Detailed MRI analyses, revealing anomalies such as cortical atrophy, ischemic changes, and ventriculomegaly, further refined the diagnostic process.

**Neurologist 4** The patient's cognitive status, ranging from normal cognition to mild cognitive impairment (MCI) or dementia, was primarily determined based on neuropsychiatric test results and the functional assessment scale. Special consideration was given to patients with Parkinson's syndrome, as their movement disorders could impact functional assessment scores. When neuropsychiatric testing clearly indicated dementia, diagnosis was straightforward. However, cases teetering on the borderline between MCI and AD required a closer examination, where functional assessment scores, medical history, and physical examination findings were collectively considered, factoring in the influence of motor disorders on the assessment. This process involved adjust-

ing the probability estimate based on clinical judgment. Regarding etiological diagnosis, a comprehensive evaluation was carried out, taking into account both medical history and imaging data. Cases presenting with Parkinson's symptoms led to differential diagnoses that included Parkinson's disease dementia, dementia with Lewy bodies, corticobasal degeneration, progressive supranuclear palsy, and others. In instances where imaging revealed markers of cerebral small vessel disease, the possibility of vascular dementia was explored. Notably, when prominent mental symptoms were coupled with significant atrophy in one side of the frontal and temporal lobes, consideration was given to frontotemporal degeneration. Infectious, metabolic, traumatic, and hereditary causes were also taken into account, guided by the relevant medical history. The adjustment of probability in these cases was guided by personal judgment.

**Neurologist 5** The assessment combined insights from clinical and medication history, specific neurological examinations, and neuropsychological test scores. Initially, attention was given to basic demographic data, such as age and the subject's living situation. Subsequently, a comprehensive evaluation of medical and social history was conducted, considering potential dementia risk factors and relevant habits. The presence or absence of ApoE alleles was noted. Medication history was scrutinized, particularly medications associated with vascular co-morbidities like antihypertensives and anticoagulants, indicative of vascular disease risk. The presence of antidepressants was acknowledged, considering potential psychiatric conditions linked to cognitive decline. During the review of neurological examinations, focus was placed on gaze, tremor, parkinsonism, and gait assessment. Neuropsychological examination scores were analyzed, first taking note of the number of abnormal tests. MoCA scores were used when available, alongside other tests like WMS. Language assessment, often relying on Animals and Digit span backwards, played a crucial role. Z scores and absolute scores were considered for test abnormality determination. Cognitive decline characterized by language and memory loss pointed to Alzheimer's disease. The presence of hallucinations and parkinsonism suggested Lewy body dementia, or if Parkinson's disease (PD) was advanced, it pointed to PD dementia. Executive dysfunction and disinhibition were signs of frontotemporal dementia. Hydrocephalus-associated urinary symptoms and specific findings hinted at normal pressure hydrocephalus. Mild cognitive impairment (MCI) was identified through mildly abnormal tests and preserved daily activities. MRIs were considered, yet clinical synopsis took precedence when imaging findings did not align with the clinical scenario. In offering a final diagnosis, a single label was assigned in cases of diagnosis confidence, while multiple labels were used if overlapping symptoms or psychiatric co-morbidities/alcoholism could obscure the presentation. In such scenarios, several labels were assigned with varying confidence levels. For instance, in equivocal cases of dementia and MCI, ratings were employed to determine the likelihood of each diagnosis. If both MCI and dementia were considered, dropdowns for each dementia subtype were used to indicate the more probable dementia type. When distinguishing between dementia and psychiatric conditions or acute encephalopathy proved challenging, all relevant options were marked alongside dementia.

**Neurologist 6** In assessing clinical cases for dementia, the process began with a comprehensive review of key demographic and historical data, encompassing details like age, gender, educational

background, family history, and existing medical comorbidities, to provide context for interpreting the cognitive presentation. The clinical records were systematically examined, with a specific focus on the critical domains relevant to diagnosing dementia syndromes. Key tools for initial assessment, such as the Mini-Mental State Exam (MMSE) and the Montreal Cognitive Assessment (MoCA) scores, provided an initial screening of the severity and pattern of cognitive impairment. Very low scores indicated advanced dementia, while higher scores within the mild impairment range prompted a more detailed review of neuropsychological test data. This battery of neurocognitive tests revealed the specific profile of cognitive deficits within domains such as memory, language, executive function, and visuospatial abilities, each of which hinted at potential etiologies. A fundamental component of the diagnostic process involved evaluating for any concurrent neurological signs, which entailed a meticulous examination of physical findings, with a particular focus on motor exam results, including assessments for rigidity, tremors, and gait disorders often associated with Parkinsonian disorders. Additionally, the Hachinski Ischemic Scale score was considered for insights into potential vascular contributions. Furthermore, it was imperative to observe the individual's functional status and any neuropsychiatric symptoms, as they bore diagnostic and prognostic significance. The clinician had to ascertain whether the deficits impeded daily activities. Behavioral manifestations such as depression, hallucinations, delusions, and agitation could provide critical distinctions between various dementia types. Once these key components were systematically reviewed, the clinician synthesized the data to formulate a comprehensive differential diagnosis. Cognitive testing profiles, behavioral presentation, family history, age of onset, and the presence of neurological signs were all weighed and considered in a holistic manner. Common differentials in dementia assessment included Alzheimer's disease, vascular cognitive impairment, dementia with Lewy bodies, Parkinson's disease dementia, and frontotemporal lobar degeneration. Lastly, the MRI results were scrutinized for any uncommon findings that could either support or contradict the differential diagnosis. This involved assessing major structural abnormalities or alterations, such as hydrocephalus or severe atrophy, which could provide further backing for the final diagnosis.

**Neurologist 7** The interpretation method followed a structured approach. Initially, cognitive impairment severity—whether normal, mild cognitive impairment (MCI), or dementia—was determined by assessing Functional Assessment Scale Score, independence level, and neuropsychiatric testing. This assessment incorporated past medical history to exclude other potential causes of functional limitations. Etiology assessment comprised several considerations. Vascular dementia was diagnosed when factors such as stroke history, cerebrovascular disease risk factors, focal neurological deficits, Hachinski infarction score, and specific MRI findings indicating infarctions, white matter hyperintensities, and perivascular spaces were present. Parkinsonism, as evaluated by the Unified Parkinson's Disease Rating Scale (UPDRS), prompted investigation for Lewy body dementia (LBD), normal pressure hydrocephalus (NPH), vascular dementia (VD), frontotemporal lobar degeneration (FTD), and variants. LBD was considered for cases with visual hallucinations, Parkinsonism, cognitive impairment, and unremarkable MRI findings, while NPH diagnosis hinged on ventricular dilation and radiological features. FTD identification relied on executive function deficits, abnormal behavior, language impairment, and MRI-documented frontal/temporal lobe atrophy. Mental illness was contemplated for individuals with relevant medical history and

substantial neuropsychiatric inventory and geriatric depression scale (GDS) symptoms. Prion disease recognition was based on distinctive MRI patterns. Conditions like infectious, metabolic, substance abuse, delirium, and psychiatric disorders were considered through medical history, coupled with the absence of specific MRI abnormalities. Lastly, multiple system atrophy (MSA) was diagnosed in cases displaying Parkinson's symptoms, defecation issues, ataxia, and cerebellar atrophy on MRI, while traumatic brain injury (TBI) diagnosis was associated with head trauma history, cognitive decline, localized lesions, and secondary atrophy.

**Neurologist 8** The evaluation process initiated with a comprehensive assessment of patient demographics, medical/family history, and risk factors. Cardiovascular and cerebrovascular risk factors were scrutinized due to their potential contribution to vascular dementia and vascular parkinsonism. Special attention was given to assessing activities of daily living (ADLs), which served as a crucial factor distinguishing dementia from mild cognitive impairment (MCI). APOE status played a pivotal role in gauging the likelihood of Alzheimer's Dementia (AD). The presence of APOE4 heightened the risk of AD, particularly in early onset cases, while APOE2 could potentially serve as a protective factor. Psychiatric history was examined to identify behavioral changes and assess whether conditions like depression or anxiety contributed to cognitive symptoms. The Geriatric Depression Scale (GDS) helped differentiate between pseudodementia/depression and other psychiatric illnesses affecting cognitive function. This information was crucial in pinpointing specific cognitive disorders (e.g., Parkinson's disease dementia, behavioral variant frontotemporal dementia (FTD), impulse control disorders in the context of dopamine agonists). A meticulous examination of clinical findings focused on gait, tremor, and bradykinesia. The presence of rest tremor, bradykinesia, or rigidity prompted consideration of parkinsonism, or other forms of parkinsonism such as Dementia with Lewy Bodies (DLB), progressive supranuclear palsy (PSP), or FTD. Comprehensive neuropsychological battery results were analyzed to discern patterns of cognitive impairment, differentiating between executive function deficits and memory impairments. Deviations in tasks such as Trails suggested executive dysfunction, potentially indicating subcortical dementia like DLB, PDD, vascular dementia, or vascular parkinsonism. Poor performance on WAIS-R or WAIS-III indicated memory impairment, typically associated with cortical dementias like AD. Imaging studies were instrumental in the evaluation. Patterns like diffuse or parietal atrophy suggested AD, while frontal-temporal atrophy indicated FTD. The presence of widespread white matter disease (WMD) burden aligned with vascular dementia or vascular parkinsonism. Specific assessments included the evaluation of the swallow tail sign, associated with PD, and midbrain atrophy, assessed through sagittal images using the midbrain-to-pons ratio (midbrain area/pontine area). Regarding the rating system, no cases received a perfect score of 100, as most presented with mixed pathologies, combining features such as amyloid beta AD changes and alpha-synuclein aggregates with parkinsonism or alpha-synuclein alongside evidence of tauopathy in PD-PSP variants. Ratings between 50% and 80% indicated varying degrees of likelihood for a specific pathology, with ratings above 80% signifying a stronger likelihood of the disease or pathology being present.

**Neurologist 9** The assessment began with a thorough review of the individual's medical history, with a focus on identifying major diagnoses that could impact cognition. This included conditions like traumatic brain injury (TBI), significant psychiatric disorders, stroke-related issues, and apoE status. Subsequently, the individual's medication history was analyzed, considering potential biases introduced by medications commonly used for Alzheimer's disease (AD) or Parkinson's disease (PD), which might have implied a higher likelihood of these conditions. Functional status assessment followed, encompassing activities of daily living (ADL) and instrumental activities of daily living (iADL), providing insights into the individual's everyday capabilities. A comprehensive physical examination was conducted, emphasizing the identification of notable abnormalities that could offer insights into cognitive status. Psychiatric and cognitive testing scales were administered, and the results were carefully analyzed for consistency and coherence. These results were also cross-referenced with the person's reported functional status. In cases of significant discrepancy, consideration was given to underlying mood or psychiatric disorders that may have influenced information accuracy. Chronology of symptoms, often absent from person-level histories, was evaluated with a particular focus on the Neuropsychiatric Inventory Questionnaire (NPI-Q), which inquired about symptoms experienced within the last 30 days. During the review of imaging studies, the gathered information was taken into account. Attention was paid to imaging findings that may have indicated AD or significant vascular disease. Unusual symptoms in the person-level history, such as new motor problems or agitation, prompted consideration of rare conditions like frontotemporal dementia (FTD), Huntington's disease, or Creutzfeldt-Jakob disease. Subsequently, a detailed review of the imaging data was conducted to identify specific features that could be indicative of these particular disorders. Lastly, the interpretation of cognitive testing scale results was influenced significantly by the individual's functional status. This guided the determination of whether the person exhibited signs of dementia, mild cognitive impairment (MCI), or fell within the spectrum of normal cognitive function. The aim was to construct a comprehensive assessment of the individual's cognitive state, accounting for these factors.

**Neurologist 10** The determination of cognitive status, including normal cognition, mild cognitive impairment (MCI), or dementia, relied primarily on neuropsychiatric test outcomes and the functional assessment scale. Notably, when individuals exhibited Parkinsonism, functional abilities were often influenced by motor impairments, making neuropsychiatric test results more influential than the Functional Activities Questionnaire (FAQ). Given the absence of distinct cutoff points for these categories, adjustments to the probability assessment were made based on individual judgment. Regarding the etiological diagnosis, a comprehensive evaluation incorporated all available clinical information and imaging data. For instance, cases presenting with Parkinsonism prompted a focused differential diagnosis that considered conditions like Dementia with Lewy Bodies (DLB), characterized by symptoms such as Parkinsonism, dementia, and hallucinations. Others included Parkinson's disease dementia (PDD), typically occurring after a prolonged history of Parkinson's disease, vascular injuries with attention to severe small vessel disease, especially within the basal ganglia, and normal pressure hydrocephalus (NPH), identified by enlarged brain ventricles. Conditions such as corticobasal degeneration (CBD) and progressive supranuclear palsy (PSP), though less common, required the presence of more typical symptoms like apraxia in CBD or abnormal vertical eye movement in PSP for diagnosis. For individuals diagnosed with MCI or

dementia but without Parkinsonism, the differential diagnosis primarily encompassed Alzheimer's disease (AD), frontotemporal dementia (FTD), and vascular injuries. FTD, for example, might exhibit pronounced non-memory impairments, along with psychiatric and behavioral symptoms, and asymmetrical brain atrophy in frontal and/or temporal lobes. Additionally, vascular injuries played a substantial role in cognitive impairment and sometimes coexisted with AD pathology. In these instances, probability assessments were adjusted based on clinical judgment. For the remaining etiologies, establishing a diagnosis necessitated a detailed clinical history.

**Neurologist 11** The evaluation process initiated with an assessment of the provided case profiles, encompassing baseline information like age, education, language, and required assistance. Supplementary data, including genetic test results such as apoe4 status, medication records, and relevant details, were also considered. Subsequently, various cognitive and physical examinations, along with associated indices, were reviewed to detect neurocognitive dysfunction. From these comprehensive case profiles, preliminary hypotheses were formulated to guide the diagnostic process, ultimately leading to specific diagnoses or a set of potential options. A meticulous evaluation of imaging studies for each case followed, examining different sequences and views for signs of cerebral atrophy or structural changes, including white matter disease. These imaging findings were correlated with case profile hypotheses to generate a list of probable diagnoses. Probability ratings were assigned to these diagnoses, reflecting the likelihood of their presence. The rating process initially involved determining whether cases met criteria for normal cognition, mild cognitive impairment (MCI), or dementia. In ambiguous cases distinguishing between dementia and MCI, probability ratings were provided for both, especially when the differentiation between MCI and mild dementia was uncertain based on testing outcomes. Subsequently, probable contributing factors to the diagnoses were identified by selecting the types of dementia most likely present. Many cases presented with multiple potential contributing causes, often including vascular dementia alongside Alzheimer's disease. Quantifying the likelihood of each diagnosis involved assigning scores of 70 or higher to those with a high probability, regardless of an individual factor's relatively low contribution to their dementia. Higher scores indicated a greater likelihood of that diagnosis being the primary cause. Causes with similar probabilities scores did not reflect an equal degree of causality to the individual's condition but merely reflected an equal probability of occurrence. Scores ranging from 20 to 30 suggested the presence of dementia, though with a minor role in the clinical presentation. Scores below 10 indicated a very low probability, implying little to no significance.

**Neurologist 12** While reviewing clinical data in conjunction with MRI scans, a notable absence was observed regarding information on symptom onset and progression. This critical aspect of history-taking has the potential to offer valuable insights into the diagnosis, as the pace of progression varies among different forms of dementia. For diagnostic purposes, reliance was placed on MMSE scores, employing a cutoff of 24 to diagnose dementia. Functional capacity assessments assisted in distinguishing between MCI and dementia. Psychiatric questionnaires proved useful in orienting toward specific diagnoses, such as Parkinson's dementia, dementia with Lewy bodies (DLB), or infectious causes. The evaluation of depression's role in cognition was challeng-

ing, but the Geriatric Depression Scale provided some guidance. In cases of uncertainty, the MRI findings played a pivotal role. For instance, clear frontotemporal atrophy with behavioral disturbances and language involvement suggested frontotemporal dementia, while temporal lobe atrophy leaned more toward Alzheimer's dementia. In cases of DLB or Parkinson's dementia, clinical presentation bore more weight when MRI results were unremarkable. Moderate to severe white matter abnormalities pointed to vascular dementia. In most cases, a shortlist of potential diagnoses was compiled before reviewing the MRI. However, there were instances where MRI results were conclusive and prompted a change in the diagnosis. For example, one case indicated possible Creutzfeldt-Jakob disease due to hallucinations and corresponding MRI findings. In another, an MRI revealed significant encephalomalacia with ventricular enlargement following a head injury. A young case with a cavum septum pellucidum was attributed to chronic traumatic encephalopathy. Lastly, global atrophy in an individual with a history of alcohol abuse and seizures pointed to alcoholic dementia. Providing a percentage of certainty for each diagnosis proved beneficial, as many cases presented mixed pathology, especially in Parkinson's dementia, where vascular disease often contributed to the clinical picture.

## Neuroradiologist approach to the ratings

**Neuroradiologist 1** The evaluation of MRI scans initiated with a global perspective to exclude multiple infarcts and identify notable brain atrophy patterns. The presence and severity of white matter lesions, chronic infarcts and microhemorrhages were recorded. Subsequent assessment focused primarily on volume loss, particularly emphasizing hemispheric asymmetry. The initial evaluation determined whether dominant frontal and anterior temporal or parietal and medial temporal volume loss was evident. A more detailed sub-analysis of each region was conducted, focusing on grading severity and documenting regional and focal volume loss in real-time. The lobar volume loss evaluation was done systematically, starting with the frontal lobes, including attention to asymmetry when present. Sub-analyses of specific regions within the frontal lobes were conducted, such as the anterior insula, cingulate gyrus, precentral gyrus, and caudate nucleus. Evaluation of temporal lobe volume loss was also carried out, distinguishing mesial and non-mesial temporal lobe atrophy. Sub-analyses of hippocampal, amygdala and parahippocampal atrophy were included, with special attention to anterior, lateral, and posterior temporal lobe atrophy, including fusiform, middle, and inferior temporal gyrus volume. The assessment for atrophy was extended to parietal and occipital lobe, documenting brainstem and cerebellar atrophy. When appraising ventricular size, a comparison was made relative to sulcal size. Findings favoring an Alzheimer's disease (AD) pattern included presence of predominant parietal and medial temporal lobe atrophy, or less frontal lobe involvement than parietal and temporal lobes. Deviations from the AD pattern, such as predominant frontal, anterior temporal, or occipital involvement, enlarged ventricles, or multiple infarcts, supported non-AD dementia patterns, including those indicative of Lewy body dementia, vascular dementia, prion disease, frontotemporal lobar degeneration and its variants, normal pressure hydrocephalus, traumatic brain injury, psychiatric diagnoses, and/or other conditions. A rating scale from 0 to 100 was employed to assess the likelihood of various diagnostic considerations. A rating of 0 was selected when no evidence supported a particular diagnosis, while a rating of 100 indicated the imaging strongly suggested that entity. Ratings of 50 were assigned when imaging findings were equally likely to represent the entity in question.

**Neuroradiologist 2** The approach to rating the cases followed a systematic checklist, starting with an assessment of the entire brain, then moving through various lobes: frontal, temporal, parietal, occipital, and the brain stem. Within this framework, the aim was to determine the possible causes of dementia based on imaging findings. Initially, features indicative of normal pressure hydrocephalus were sought. These features typically stood out from other conditions and included disproportionate ventricular enlargement, an acute callosal angle at the posterior commissure level, sulcal crowding near the vertex, and Sylvian fissure enlargement. Next, the focus shifted to assessing the overall burden of white matter disease, characterized by T2 FLAIR hyperintensities. Examination was carried out in regions with encephalomalacia or gliosis, which might signify prior infarcts, helping establish a potential vascular component to dementia, either as the sole cause or a contributing factor alongside other processes. Further examination was directed toward atrophy patterns, aiming to identify specific neurodegenerative processes. Disproportionate atrophy in the medial, basal, and lateral temporal lobes and the medial parietal lobes suggested Alzheimer's

disease. Relative preservation of medial temporal lobe structures hinted at dementia with Lewy bodies or Parkinson's disease dementia, although the absence of clinical history posed challenges for this diagnosis, as clinical features and typical MRI findings of medial temporal lobe preservation are valuable in a clinical setting. For frontotemporal lobar degeneration and its variants, the search was for frontal and/or temporal atrophy, predominately left posterior perisylvian or parietal atrophy, anterior temporal atrophy, predominant left posterior fronto-insular atrophy, midbrain atrophy relative to the pons ('hummingbird' sign), concavity of the dorsolateral midbrain, thinning of the tectal plate, or T2 hyperintense rim along the putamen with patchy or confluent T2 FLAIR hyperintensity in the rolandic subcortical white matter. In the quest for Prion disease indicators, examination included cortical/gyriform diffusion hyperintensity, often accompanied by thalamic and basal ganglia diffusion hyperintensity. Also explored were signs of encephalomalacia and gliosis typical of prior traumatic brain injury.

**Neuroradiologist 3** During case reviews, emphasis was placed on patient age and MRI findings as essential factors guiding the diagnostic process. Age served as a key determinant, informing the assessment of volume loss, particularly relevant in cases of Alzheimer's and frontotemporal lobar degeneration (FTLD). Each MRI sequence contributed uniquely to diagnostic considerations: T1-weighted images held importance in gauging volume loss, discerning distinctive patterns within the hippocampus, temporal lobes, and parietal lobes for Alzheimer's disease, and focusing on volume loss within the frontal and temporal lobes for FTLD. In the assessment for normal pressure hydrocephalus, attention was drawn to ventriculomegaly and its proportionality to volume loss. T1-weighted images were also instrumental in identifying cerebellar atrophy, indicative of conditions like alcoholism or phenytoin use for seizures. Diffusion-weighted images played a critical role in detecting signs of Creutzfeldt-Jakob Disease, characterized by hyperintensity in regions such as the insula, cingulate gyrus, frontal gyri, medial thalami, and possibly the basal ganglia. This sequence was also valuable for identifying infarcts. T2/FLAIR and other T2-weighted images were essential for assessing small vessel disease burden, aiding in the evaluation of vascular dementia. They were also instrumental in detecting potential evidence of infectious, inflammatory, metabolic, or drug-related hyperintensity. The susceptibility-weighted images were used to assess for microhemorrhages, which could be associated with Alzheimer's or Lewy body disease. Psychiatric diseases were typically exempt from numerical ratings as their diagnosis could not usually be ascertained through imaging. Ratings spanned from 70 to 90 in cases where a single diagnosis was highly confident. In scenarios where multiple potential diagnoses were considered, ratings ranged from 40 to 70 for each disease state, reflecting the estimated likelihood of each condition.

**Neuroradiologist 4** Each case was approached by first reviewing the demographic information; however, as the project progressed, the demographic data became less informative, and by the midpoint of the project, demographics were reviewed only as a later step. The images were assessed using the SLICER software. The T2 and FLAIR sequences were carefully evaluated to gauge the extent of small vessel disease and infarcts, serving as indicators of potential vascular causes of cognitive impairment. These sequences also proved valuable for the exclusion of infectious, inflammatory, or toxic causes. The DWI sequence was employed to identify acute infarcts and

to investigate neurodegenerative conditions such as Creutzfeldt-Jakob disease or fatal familial insomnia. Susceptibility-weighted images were analyzed to identify microhemorrhages, assess their extent and location, and rule out other potential causes of cognitive decline. However, the most pivotal sequences were the volumetric sequences acquired in all three anatomical planes. They were instrumental in assessing global or lobar-specific volume loss. Specific regions of interest included the hippocampal volume assessed through coronal sequences to rule out Alzheimer's disease, the precuneus evaluated via sagittal sequences, and the parietal lobes examined in axial sequences. If frontal lobe volume loss was evident, then the temporal lobes were assessed for signs of frontotemporal dementia. Cerebellar volume loss or infratentorial volume loss led to considerations of alcohol abuse or phenytoin use, or cerebellar ataxias, while brainstem involvement indicated potential multisystem atrophy. Disproportionate ventricular dilatation raised suspicions of normal pressure hydrocephalus. The rating scale used was comprehensive, and in cases where complete information was lacking, the diagnosis was assigned to the best of the ability. A diagnosis was rated as 100 when highly confident, and as 50 when uncertainty existed. Additionally, some cases were assigned a probability score between 50 and 100 when confident in excluding other potential causes, based on the imaging data.

**Neuroradiologist 5** The approach to MR exams began with an evaluation of axial T2/FLAIR images, if available. If multiple regions of gliosis were observed alongside areas of encephalomalacia, resulting from prior infarctions in multiple vascular territories, consideration was given to the possibility of multi-infarct dementia. Moreover, when encephalomalacia and gliosis predominantly affected the temporal lobes, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy became a potential inclusion in the diagnostic considerations. Following the FLAIR sequence, assessment of diffusion-weighted images, if accessible, primarily served to rule out more acute conditions like Creutzfeldt-Jakob disease, herpes encephalitis, or other forms of encephalitis. Subsequently, T1-weighted images were reviewed, preferably in 3D format, to examine ventricle and sulci dimensions. The presence of ventriculomegaly and sulcal crowding at the vertex prompted consideration of normal pressure hydrocephalus as a potential diagnosis. Additionally, gyri were evaluated to identify areas exhibiting volume loss. T2-weighted images were especially helpful in this regard, as they enhanced the visibility of cerebrospinal fluid and accentuated regions of atrophy. Once the order of diagnostic differentials was established, a diagnostic rating was assigned. In this rating system, a score of 100 indicated absolute certainty, an exceedingly rare occurrence in radiology. Conversely, a score of less than 20 signified extreme unlikelihood, 25 denoted unlikeliness, 50 implied the possibility of the diagnosis, while a range of 50-75 indicated a probable diagnosis. Finally, a score exceeding 75 suggested a high likelihood of the diagnosis being accurate.

**Neuroradiologist 6** The review process began with an examination of the provided individual-level demographics for each case. Subsequently, all images provided for each case underwent analysis using the SLICER software. T2/FLAIR sequence was the basis for assessing small vessel changes, subacute to chronic infarcts, encephalomalacia from traumatic brain injury, and any areas displaying signal abnormalities indicative of potential alternative causes, such as neurodegener-

ative, infectious-inflammatory, or toxic-metabolic etiologies. T2/FLAIR sequence was also employed to investigate seizure-related changes. T2-weighted images played a key role in evaluating ventricular size, examining the posterior fossa for small infarcts, and observing major intracranial arterial flow voids. Diffusion-weighted images were used to identify acute infarcts and regions with reduced diffusivity, potentially linked to other neurodegenerative, infectious-inflammatory, toxic-metabolic conditions, or seizure-related changes. Susceptibility-weighted images were utilized to detect areas featuring parenchymal microhemorrhage or calcification. Lastly, high-resolution T1-weighted images were employed to analyze regional volume loss patterns suggestive of specific neurodegenerative processes. The evaluation process included the completion of the online ADRD radiologist task survey. During the assessment of sections regarding regional predominate atrophy, the high-resolution T1-weighted images were revisited to ensure response accuracy. In the final section, person-level demographics and imaging findings were synthesized to arrive at the best-guess probability for each diagnosis. The rating scale corresponded to the likelihood of the best-guess diagnosis. For instance, if there was high confidence that a case represented a particular diagnosis, it was assigned a score of 100, with a score of 0 given to all other diagnoses. In cases of diagnostic uncertainty, where the estimated probability was 50%, a score of 50 was assigned.

**Neuroradiologist 7** Brain volume loss was assessed based on age-appropriate norms, with T1 and T2/FLAIR sequences aiding in the evaluation of volume loss within each lobe. These sequences were particularly useful for assessing cerebrospinal fluid presence near the convexity. Brainstem volume loss was primarily evaluated through mid-sagittal and axial images, which allowed for the examination of the pontine belly and cerebral peduncle size, respectively. Coronal images provided insights into hippocampal volume, determined by the prominence of the temporal horns of the lateral ventricle. Sagittal images were employed to assess cerebellar volume loss. FLAIR sequences played a crucial role in detecting encephalomalacia, gliosis, infarcts, and white matter changes. Distinct patterns were observed in various dementia types, such as parietotemporal volume loss favoring Alzheimer's dementia. Extensive white matter changes with or without microhemorrhages in individuals over 60 years pointed to vascular dementia. White matter changes in younger individuals raised consideration of alternative causes like infections or metabolic factors. Alcohol use often correlated with cerebellar volume loss. Traumatic brain injury was suspected in cases with FLAIR signal changes and peripheral volume loss in the anterior temporal and inferior frontal lobes, with or without susceptibility, along with corpus callosum and brainstem findings, suggestive of diffuse axonal injury. Frontal and temporal lobe volume loss indicated frontotemporal dementia. The 'hummingbird' sign on sagittal images led to consideration of progressive supranuclear palsy, particularly when combined with brainstem volume loss. Asymmetric ventricular prominence relative to cortical volume loss hinted at normal pressure hydrocephalus, with the corpus callosal angle measured on coronal images to confirm the diagnosis. While no specific findings were linked to psychiatric disorders, the presence of a cavum septum pellucidum was weakly correlated. Multiple findings in a case, such as global volume loss, extensive white matter changes and microhemorrhages, leaned toward vascular dementia over Alzheimer's disease due to the subjective nature of volume loss assessment. A higher rating was assigned to the diagnosis with more MRI findings supporting it, though no case received a perfect score of 100, with ratings exceeding 80 indicating a dominant diagnosis.