



Diagnostic Assessment & Prognosis

Evaluating combinations of diagnostic tests to discriminate different dementia types

Marie Bruun^{a,*¹}, Hanneke F. M. Rhodius-Meester^{b,1}, Juha Koikkalainen^c, Marta Baroni^d, Le Gjerum^a, Afina W. Lemstra^b, Frederik Barkhof^{e,f}, Anne M. Remes^{g,h}, Timo Urhemaaⁱ, Antti Tolonenⁱ, Daniel Rueckert^j, Mark van Gilsⁱ, Kristian S. Frederiksen^a, Gunhild Waldemar^a, Philip Scheltens^b, Patrizia Mecocci^d, Hilkka Soininen^k, Jyrki Lötjönen^c, Steen G. Hasselbalch^a, Wiesje M. van der Flier^{b,1}

^aDanish Dementia Research Centre, Department of Neurology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

^bAlzheimer Center, Department of Neurology, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, the Netherlands

^cCombinostics Ltd., Tampere, Finland

^dInstitute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy

^eDepartment of Radiology and Nuclear Medicine, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands

^fUCL Institutes of Neurology and Healthcare Engineering, London, United Kingdom

^gMedical Research Center, Oulu University Hospital, Oulu, Finland

^hUnit of Clinical Neuroscience, Neurology, University of Oulu, Oulu, Finland

ⁱVTT Technical Research Center of Finland Ltd, Tampere, Finland

^jDepartment of Computing, Imperial College, London, United Kingdom

^kInstitute of Clinical Medicine, Neurology, University of Eastern Finland, Kuopio, Finland

^lDepartment of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, the Netherlands

Abstract

Introduction: We studied, using a data-driven approach, how different combinations of diagnostic tests contribute to the differential diagnosis of dementia.

Methods: In this multicenter study, we included 356 patients with Alzheimer's disease, 87 frontotemporal dementia, 61 dementia with Lewy bodies, 38 vascular dementia, and 302 controls. We used a classifier to assess accuracy for individual performance and combinations of cognitive tests, cerebrospinal fluid biomarkers, and automated magnetic resonance imaging features for pairwise differentiation between dementia types.

Results: Cognitive tests had good performance in separating any type of dementia from controls. Cerebrospinal fluid optimally contributed to identifying Alzheimer's disease, whereas magnetic resonance imaging features aided in separating vascular dementia, dementia with Lewy bodies, and frontotemporal dementia. Combining diagnostic tests increased the accuracy, with balanced accuracies ranging from 78% to 97%.

Discussion: Different diagnostic tests have their distinct roles in differential diagnostics of dementias. Our results indicate that combining different diagnostic tests may increase the accuracy further.

© 2018 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords:

Differential diagnosis; Biomarkers; Diagnostic test assessment; Clinical decision support system; CSF; MRI; Alzheimer's disease; Frontotemporal dementia; Dementia with Lewy bodies; Vascular dementia

The authors have declared that no conflict of interest exists.

¹Both authors contributed equally to the article and share first authorship.

*Corresponding author. Tel.: +45 35457662, Mobile: +45 30240724; Fax: +45 35452446.

E-mail address: marie.bruun@regionh.dk

1. Background

Dementia affects an increasing number of people worldwide [1]. Alzheimer's disease (AD) is the most frequent cause of dementia accounting for 50%–70% of dementia

cases [2]. Other common causes of dementia include vascular dementia (VaD), dementia with Lewy bodies (DLB), and frontotemporal dementia (FTD) [3–6]. To ensure appropriate pharmacological treatment, counseling, and inclusion in clinical trials, early and precise diagnosis of the underlying disease causing dementia is important.

Cognitive profiles differ between dementia types showing primarily memory impairment in AD, visuospatial and executive dysfunction in DLB, delayed cognitive processing in VaD, and mainly language, executive, and behavioral dysfunction in FTD, although considerable overlap exists [7,8]. Progress in biomarker development has provided new disease insights and improved accuracy of dementia diagnosis. This has led to an increasing role of biomarkers, such as those obtained from cerebrospinal fluid (CSF) measures and structural magnetic resonance imaging (MRI), in diagnostic criteria and guidelines [3–6]. CSF biomarkers can provide evidence for the presence of β amyloid 1–42 ($A\beta_{42}$) accumulation and downstream neuronal dementia in AD (tau and tau phosphorylated at threonine 181 [p-tau]), whereas isolated elevation of tau may also be seen in FTD, and intermediate concentrations of CSF biomarkers often occur in DLB and VaD [9,10]. On structural MRI, typical abnormalities for different causes of dementia have been described, such as hippocampal and parietal atrophy in AD, frontal-temporal atrophy in FTD, and profound white matter hyperintensities (WMHs) in VaD, whereas DLB presents with unspecific mild generalized atrophy [11–13].

Despite these advances, differential diagnosis of dementia in terms of accurately identifying the underlying etiology remains challenging. First, biomarkers for other types of dementia are less developed than those for AD; and second, there is often overlap in underlying pathology and clinical presentation as most patients do not present in an archetypical fashion [9,11]. In addition, diagnostic guidelines remain relatively general and address one disease only. In reality, a clinician is often faced with a complex differential diagnostic task of simultaneously evaluating a range of potential diagnoses and combining data from multiple tests and biomarkers. More knowledge on performance and value of biomarkers in the differential diagnosis of dementia is therefore needed.

Combination of diagnostic tests, such as MRI and CSF, has been studied for AD and progression, but not previously for differential diagnosis in a multicenter cohort [14,15]. We used a classifier based on the Disease State Index (DSI) [16] in a large cohort from four European memory clinics to differentiate between controls and patients with AD, FTD, DLB, and VaD. We used a data-driven approach to explore the diagnostic accuracy of commonly used clinical diagnostic tests, including cognitive tests, CSF biomarkers, and automated MRI features. Furthermore, we evaluated performance of all diagnostic tests combined and which combinations of tests were optimal for each pairwise comparison of diagnoses.

2. Methods

2.1. Subjects

We included 844 subjects, which were pooled from four different memory clinic-based cohorts: 543 subjects from the Amsterdam Dementia Cohort at the VU Medical Center Amsterdam [17,18], 112 subjects from the Danish Dementia Research Center at Copenhagen University Hospital, Rigshospitalet, 139 subjects from the Department of Gerontology and Geriatrics of the University of Perugia, “S. Maria della Misericordia” Hospital of Perugia, and 50 subjects from the Department of Neurology from the University of Eastern Finland. Data from Rigshospitalet, University of Perugia, University of Eastern Finland, and 44 subjects from VU Medical Center had been collected as part of the PredictND study [19]. The remainder of VUmc subjects was included from Amsterdam Dementia Cohort. The pooled cohort consisted of subjects with the following diagnosis: 326 AD, 87 FTD, 61 DLB, 38 VaD, and 302 controls with subjective cognitive decline (SCD) (Table 1). Subjects were eligible for inclusion if brain MRI was available.

All subjects had received a standardized workup, including medical history, physical, neurological and neuropsychological assessment, MRI, laboratory tests, and a subset examination of CSF. Individuals were diagnosed as SCD when the cognitive complaints could not be confirmed by cognitive testing and criteria for mild cognitive impairment or dementia were not met. The diagnoses were established based on the following diagnostic criteria: the criteria of the NIA-AA for AD dementia [3], the Rascovsky and Gorno-Tempini criteria for FTD [5,20], the NINDS-AIREN criteria for VaD [4], and the McKeith criteria for DLB [6,21]. All patients had provided written informed consent for their data to be used for research purposes.

2.2. Clinical assessment

2.2.1. Neuropsychology

We used the Mini-Mental State Examination for global cognitive functioning [22]. For memory, the Consortium to Establish a Registry for Alzheimer's Disease word list memory test and the Rey Auditory Verbal Learning Task were included [23,24]. To measure cognitive processing speed and executive functioning, we used Trail Making Tests A and B (TMT-A and TMT-B, respectively) [25]. Language and executive functioning were tested by category fluency (animals) [26]. Missing data ranged from $n = 1$ (Mini-Mental State Examination) to $n = 31$ (4%) (memory). To pool the different memory tests, we standardized Rey Auditory Verbal Learning Task and Consortium to Establish a Registry for Alzheimer's Disease memory tests per center to z-scores using SCD subjects.

2.2.2. Cerebrospinal fluid biomarkers

The CSF biomarkers $A\beta_{42}$, total tau, and p-tau were measured with commercially available ELISA tests

Table 1
Baseline characteristics according to baseline diagnosis

	n	Control, n = 302	AD, n = 356	FTD, n = 87	VaD, n = 38	DLB, n = 61	Group-wise comparisons when significant
Female, n (%)	844	168 (56)	201 (56)	39 (45)	14 (37)	9 (15)	Controls, AD < FTD, VaD < DLB
Age, in years	844	63 ± 9	68 ± 8	63 ± 7	72 ± 8	69 ± 8	Controls, FTD < AD, DLB, VaD
Cognitive tests							
MMSE	840	29 ± 1	22 ± 4	24 ± 5	24 ± 4	23 ± 4	Controls > FTD, VaD, DLB > AD
Memory, learning	803	44 ± 10	23 ± 9	27 ± 8	25 ± 8	25 ± 9	Controls > FTD > AD, VaD, DLB
Memory, recall	803	9 ± 3	2 ± 2	4 ± 3	3 ± 3	4 ± 3	Controls > FTD, VaD, DLB > AD
TMT-A, in seconds	843	38 ± 17	93 ± 76	70 ± 58	112 ± 82	115 ± 81	Controls < FTD < AD, VaD, DLB
TMT-B, in seconds	829	86 ± 42	228 ± 84	188 ± 93	254 ± 69	261 ± 73	Controls < FTD < AD, VaD, DLB
Animal fluency	815	24 ± 7	13 ± 5	12 ± 7	11 ± 4	13 ± 5	Controls > AD, FTD, VaD, DLB
CSF							
Aβ ₄₂ , pg/mL	596	922 ± 274	530 ± 167	918 ± 252	727 ± 264	741 ± 264	Controls, FTD > VaD, DLB > AD
Total tau, pg/mL	582	307 ± 178	690 ± 407	330 ± 126	314 ± 162	344 ± 222	Controls, FTD, VaD, DLB < AD
p-tau, pg/mL	592	50 ± 21	85 ± 39	44 ± 18	44 ± 18	51 ± 28	Controls, FTD, VaD, DLB < AD
MRI							
Hippocampus, mL	844	7.0 ± 0.8	5.7 ± 0.9	5.7 ± 1.0	6.0 ± 1.0	6.3 ± 0.9	Controls > DLB > AD, FTD, DBL
Lateral ventricle, mL	844	29 ± 15	49 ± 20	46 ± 20	61 ± 28	43 ± 15	Controls < AD, FTD, DLB < VaD
Inferior lateral ventricle, mL	844	1.3 ± 0.5	2.5 ± 1.1	3.1 ± 2.0	3.0 ± 1.5	2.0 ± 0.9	Controls < DLB < AD, FTD, VaD
Cortex, mL	844	510 ± 34	462 ± 29	462 ± 35	453 ± 29	471 ± 22	Controls > AD, FTD, DLB > VaD
Frontal cortex, mL	844	201 ± 15	185 ± 13	179 ± 19	177 ± 14	189 ± 10	Controls > AD, DLB > FTD, VaD
Temporal cortex, mL	844	125 ± 9	109 ± 10	105 ± 15	114 ± 10	115 ± 7	Controls > VaD, DLB > AD, FTD
Medial temporal cortex, mL	844	20 ± 2	16 ± 2	16 ± 3	18 ± 3	18 ± 2	Controls > VaD, DLB > AD, FTD
Parietal cortex, mL	844	111 ± 9	99 ± 8	106 ± 9	97 ± 7	101 ± 6	Controls > FTD > AD, VaD, DLB
Occipital cortex, mL	844	74 ± 7	68 ± 7	72 ± 7	66 ± 8	67 ± 6	Controls, FTD > AD, VaD, DLB
Anterior versus posterior index	844	-0.3 ± 0.8	-0.0 ± 1.1	-2.1 ± 1.7	-0.3 ± 1.0	0.0 ± 0.8	Controls, AD, VaD, DLB > FTD
WMH, mL	844	2.8 ± 4.7	6.1 ± 9.2	3.8 ± 8.7	34.9 ± 26.7	4.4 ± 6.0	Controls < AD, FTD, DLB < VaD
Cortical infarcts, mL	844	0.0 ± 0.4	0.1 ± 2.1	0.3 ± 1.8	4.1 ± 6.8	0.0 ± 0.0	Controls, AD, FTD, DLB < VaD
Lacunar infarcts, mL	844	0.0 ± 0.0	0.0 ± 0.1	0.0 ± 0.0	0.2 ± 0.2	0.0 ± 0.0	Controls, AD, FTD, DLB < VaD

Abbreviations: AD, Alzheimer's disease; FTD, frontotemporal dementia; VAD, vascular dementia; DBL, dementia with Lewy bodies; MMSE, Mini-Mental State Examination; TMT, Trail Making Test; Aβ₄₂, β amyloid 1–42; p-tau, tau phosphorylated at threonine 181; WMH, white matter hyperintensity; MRI, magnetic resonance imaging; RAVLT, Rey Auditory Verbal Learning Task; CERAD, Consortium to Establish a Registry for Alzheimer's Disease.

NOTE. Data are presented as mean ± SD, unless otherwise specified. Group differences were calculated using one-way ANOVA with post hoc Bonferroni for continuous variables. For categorical variables, χ-square test was used. P < .05 = significant.

NOTE. Memory: RAVLT values, using z-scoring for those with only CERAD. MRI: volumes are defined from image segmentations produced by a multi-atlas segmentation algorithm, and we report the sum of left and right, in mL.

NOTE. Voxel- and tensor-based morphometry features are computed separately for each disease pairs and reported in Appendix; also ROI-based grading consists of eight features, so reported in the [Supplementary Appendix](#). MRI volumes are adjusted for head size.

(Innotest, Fujirebio, Ghent, Belgium) locally according to standard procedures. Raw data values were used for analysis. Analysis showed comparable values between the centers, and no further correction were performed (see [Supplementary Fig. 3A in Appendix](#)). CSF data were available for 596 (71%) of the subjects.

2.2.3. Imaging biomarkers

MRI scans were acquired on 1.5 T or 3 T scanners with slice thickness of 3D T1-weighted images <1.5 mm. MRI scans were available for all subjects. Imaging biomarkers were extracted from T1-weighted and fluid-attenuated inversion recovery images using image quantification methods, described in detail in [27].

Volumes of 133 brain regions were defined from image segmentations produced by a multi-atlas segmentation algorithm [28]. We used the volumes of the following 10 clinically relevant regions as imaging biomarkers: hippocampus, inferior lateral ventricle, lateral ventricle, frontal

cortex, medial temporal cortex, temporal cortex, temporal pole, parietal cortex, occipital cortex, and whole cortex. Furthermore, we derived from the volumes an imaging biomarker called anterior versus posterior index, which is defined as a ratio of the volumes at frontal and temporal lobe regions to the volumes at parietal and occipital lobe regions [29].

In addition to volumetry, three other quantification methods were applied to T1 images: voxel-based morphometry (VBM), tensor-based morphometry (TBM), and automated region-of-interest (ROI)-based grading. VBM measures the local concentration of gray matter [30], and TBM measures the local volume differences between images [31]. For each pair of diagnostic groups, we defined a pattern of locations where significant differences are found in the local concentration or volume. VBM and TBM indices used in this work measure how well a patient fits to these patterns. The ROI-based grading reconstructs a certain ROI of the patient image as a weighted linear combination of the

corresponding ROIs from a library of reference images. In addition to the image itself, each reference image contains information about the subject's diagnostic label. The grading feature is defined as the share of the weights from the linear model having a certain diagnostic label. Eight grading features that describe the similarities to SCD, AD, FTD, and DLB data set images in the hippocampus and frontal lobe regions were computed [32]. Finally, vascular burden was measured with three features from fluid-attenuated inversion recovery images: volume of WMHs, volume of cortical infarcts, and volume of lacunar infarcts [27]. Additional technical information is available in Appendix.

2.3. Disease State Index

The DSI classifier used in this study is composed of two components: fitness and relevance [33]. Fitness of diagnostic test i , as a function of diagnostic test value x , is defined as

$$f_i(x) = \frac{FN_i(x)}{FN_i(x) + FP_i(x)},$$

where FN is the false negative rate and FP is the false positive rate if the value x is used to classify the diagnostic test. A fitness value of 0 indicates a high similarity to the negative group, for example, controls (or SCD), whereas a fitness value of 1 indicates a high similarity to the positive group, for example, AD. Relevance of a diagnostic test i defines how good that diagnostic test (combination) is in differentiating the negative and positive groups. Relevance is defined as *sensitivity* + *specificity* - 1. Finally, DSI is computed for multiple diagnostic tests i as

$$DSI = \frac{\sum_i \text{relevance}_i \times \text{fitness}_i}{\sum_i \text{relevance}_i}$$

The DSI is a scalar value between 0 and 1 describing the similarity of the patient's data to comparison of each pair of diagnostic groups in the database [16,33]. In classification, the cutoff value is 0.5. The DSI classifier is an ensemble classifier, which by definition is able to deal with missing data and no imputation is needed.

2.4. Data analysis

Baseline characteristics were compared using parametric and nonparametric tests where appropriate. MRI volume biomarkers were corrected for the head size [34], age, and sex [35]. Remaining markers were corrected for age and sex [35]. We assessed the performance of each individual diagnostic test in differentiating between diagnostic groups through pairwise comparison. Subsequently, we assessed the performance when combining the three groups of tests (cognitive tests, CSF biomarkers, and MRI features) in different compositions. We used the following performance metrics: balanced accuracy and area under the receiver operating characteristic curve (AUC). For two diagnostic groups,

balanced accuracy can be defined simply as the average of sensitivity and specificity for one threshold and addresses the imbalance in the number of cases between different diagnostic groups. The AUC provides the percentage of area under the receiver operating characteristic curve representing true positive rates versus false positive rates across all possible thresholds. All performance measures were computed using 10-fold cross-validation. The disease-specific patterns used in the VBM and TBM indices (see Section 2.2.3) were defined within this same cross-validation procedure.

Feature selection is often applied because certain diagnostic tests can become redundant or irrelevant when using multiple tests. In DSI, less-relevant diagnostic tests have small weights, but they are not totally excluded from computations as happens in feature selection. Therefore, we applied finally a separate feature selection step to see which tests become selected in the optimal combination. Diagnostic tests were added one-by-one if the AUC increased. When the optimal sets of diagnostic tests are searched using cross-validation, each fold leads to a slightly different set of tests. Therefore, we applied feature selection to all data without cross-validation.

$P < .05$ was considered significant. Statistical analyses were performed using SPSS, version 22 (IBM, Armonk, NY). A MATLAB toolbox created by Cluitmans et al. [36] was used in the DSI analyses. The analyses were performed in MATLAB, version R2015b (MathWorks, Natick, MA).

3. Results

3.1. Subjects

Baseline characteristics according to diagnostic groups are presented in Table 1. Based on raw data comparison, controls performed better on neuropsychological testing and showed less atrophy on MRI features than all types of dementia. Subjects with AD scored lowest on memory tests and had lower A β ₄₂ and higher tau CSF biomarkers. Subjects with VaD were older, showed most vascular changes in MRI features (WMHs, and cortical and lacunar infarcts), and, together with subjects with DLB, performed slower on TMT-A and TMT-B. Subjects with FTD had lower anterior versus posterior index values.

Details on VBM and TBM features (computed separately for each disease pairs) and ROI-based grading (consisting of eight features) can be found in Appendix (Supplementary Table 1A).

3.2. Accuracy of each diagnostic test for differentiation of diagnostic groups

Table 2 provides a detailed view of the performance of each individual diagnostic test for each pairwise comparison of diagnostic groups showing balanced accuracy and AUC. Cognitive tests and automated MRI features showed high accuracy for separating controls from patients with

Table 2

Pairwise comparison of diagnostic groups for all diagnostic tests, reporting balanced accuracy and area under the ROC curve (Bal. Acc./AUC)

	Controls	Controls	Controls	Controls	AD	AD	AD	FTD	FTD	VaD
	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.	vs.
	AD	FTD	VAD	DLB	FTD	VaD	DLB	VAD	DLB	DLB
Cognitive tests										
MMSE	91/97	82/89	83/91	88/94	60/65	53/60	54/61	55/55	55/55	46/44
Memory, learning	87/94	83/90	86/91	83/90	61/63	60/60	54/57	43/45	56/56	55/52
Memory, recall	89/95	81/88	80/88	75/85	67/68	63/68	72/76	47/46	52/57	58/59
TMT-A	75/82	71/77	84/89	83/91	55/58	57/60	57/62	63/68	66/71	47/52
TMT-B	82/91	76/84	89/94	90/94	58/59	43/46	53/52	61/60	64/61	36/39
Animal fluency	82/90	77/88	89/96	79/88	55/57	60/62	56/51	53/52	54/56	58/61
CSF										
AB42	83/88	50/49	58/67	63/67	82/90	69/73	67/75	58/67	61/67	50/45
Total tau	75/83	58/59	40/42	54/51	75/80	76/83	77/81	63/62	58/57	57/53
P-tau	70/79	59/60	59/62	56/54	77/84	75/85	71/78	49/48	46/47	53/58
MRI										
Hippocampus	77/86	79/85	70/74	64/67	50/48	62/65	69/73	64/65	72/74	56/59
Lateral ventricle	75/81	74/81	76/84	68/73	46/45	58/58	56/60	58/57	57/61	65/68
Inferior lateral ventricle	79/85	78/87	79/88	64/69	59/60	50/51	61/68	50/54	67/74	63/72
Temporal pole	72/80	84/89	68/70	64/68	68/74	55/60	60/66	73/79	76/82	59/57
Cortex	77/85	79/85	81/84	71/76	56/57	43/42	61/63	54/55	65/68	62/64
Frontal cortex	69/76	77/83	76/81	65/67	59/65	60/62	57/61	54/57	66/72	67/70
Temporal cortex	81/87	82/87	71/74	73/79	58/61	64/67	63/69	66/72	66/74	48/47
Medial temporal cortex	81/89	82/88	66/72	66/74	58/61	66/69	67/73	71/74	74/77	41/37
Parietal cortex	72/80	61/66	74/82	64/73	61/64	43/38	57/60	56/65	40/46	57/61
Occipital cortex	62/67	57/60	62/70	65/70	57/58	51/54	59/57	58/62	61/65	52/48
Anterior vs. Posterior index	54/56	77/84	38/35	56/61	79/85	54/56	52/53	72/82	81/88	61/60
Global VBM-index	82/89	84/91	88/92	77/84	65/71	52/48	67/73	59/65	77/84	60/69
Global TBM-index	80/87	85/90	87/91	73/81	70/77	59/64	57/61	82/88	72/80	61/70
ROI-based grading	83/91	81/91	83/88	73/80	68/76	54/62	65/71	66/70	73/78	64/67
Vascular burden	61/64	55/54	87/94	59/58	55/58	84/90	48/49	86/91	44/44	85/92

Abbreviations: ROC, receiver operating characteristic; AD, Alzheimer's disease; FTD, frontotemporal dementia; VaD, vascular dementia; DLB, dementia with Lewy bodies; MMSE, Mini-Mental State Examination; TMT, Trail Making Test; A β 42, β amyloid 1–42; p-tau, tau phosphorylated at threonine 181; VBM, voxel-based morphometry; TBM, tensor-based morphometry; ROI, Region of interest; WMH, white matter hyperintensity; AUC, area under the receiver operating characteristic curve; Bal.Acc., balanced accuracy; ROC, receiver operating characteristic; RAVLT, Rey Auditory Verbal Learning Task; CERAD, Consortium to Establish a Registry for Alzheimer's Disease.

NOTE. For each pairwise comparison of two diagnostic groups, a balanced accuracy and AUC (Bal.Acc./AUC) are presented for each diagnostic test. Balanced accuracies 85%–100% are highlighted in dark green. The gradually lighter shades of green indicate lower balanced accuracy with white being at or below 50. Both balanced accuracy and AUC are reported as percentage values (%).

NOTE. Memory: RAVLT values, using z-scoring for those with only CERAD. MRI: volumes are defined from image segmentations produced by a multi-atlas segmentation algorithm. Anterior versus posterior index: index between anterior and posterior weighted volumes. ROI-based grading: based on hippocampus region of interest. The classification for "grading" consists of eight grading features and "vascular burden" consists of three features: volume of WMHs, volume of cortical infarcts, and volume of lacunar infarcts.

any type of dementia. However, cognitive tests had low accuracy in differentiating between types of dementia. CSF biomarkers had the highest accuracy for differentiating AD from other types of dementia, and especially, A β ₄₂ had high separating performance for AD vs. FTD. Vascular burden features were the most important diagnostic tests to identify VaD, also in the AD vs. VaD comparison. Automated MRI quantification features had the highest performance for separating FTD and DLB from other diagnostic groups. Especially, the medial temporal cortex volume, anterior versus posterior index, and ROI-based grading feature supported differentiation of FTD and DLB from controls, AD, and VaD, but also the pairwise comparison FTD vs. DLB. The anterior versus posterior index performed almost as well as A β ₄₂ for separating AD and FTD and had the highest accuracy in the FTD vs. DLB comparison. When we repeated the analysis in a subset with complete data only, results were comparable (data not shown).

3.3. Combinations of diagnostic tests for differentiation between diagnostic groups

Table 3 shows, for each pairwise comparison of diagnostic groups, the balanced accuracies and AUC for different combinations of the three groups of tests (cognitive tests, CSF, and MRI), using all tests without a specific feature selection. The best group of tests for differentiation of AD was CSF biomarkers. Combined MRI features generally achieved a high performance, especially in pairwise comparison including VaD or FTD. Except for pairwise compar-

sions including VaD and FTD vs. DLB, accuracy increased for all diagnostic pairs compared with the best single diagnostic tests in **Table 2** when combining all tests, ranging from 78% (for VaD vs. DLB) to 97% (for controls vs. AD).

Table 4 presents the optimal sets of diagnostic tests for each pairwise diagnostic comparison using feature selection. The results show that each of the optimal set of tests was different. Overall, the optimized combinations separating controls from dementia were a composition of cognitive tests and MRI. For AD vs. FTD, the inclusion of only CSF biomarkers provided the highest accuracy, whereas for separation of FTD vs. VaD and VaD vs. DLB, only MRI features were included. For AD vs. VaD, AD vs. DLB, and FTD vs. DLB, different combinations of cognitive tests, CSF biomarkers, and MRI features provided the best performance.

4. Discussion

In this multicenter study, we determined the diagnostic performance of cognitive tests, CSF biomarkers, and automated MRI features in differential diagnosis of dementia using a data-driven approach in a multicenter memory clinic cohort. Our results show the performance of individual diagnostic tests differed between pairwise comparisons of diagnostic groups. Moreover, combining diagnostic tests improved accuracy for pairwise diagnostic comparison.

We used the DSI classifier that can deal with missing data and make pairwise comparisons [16]. Previously, we have shown this classifier capable of separating different types of dementia with high accuracy using retrospective data from one memory clinic [37]. We extended this work by

Table 3

Performance of diagnostic tests for each pairwise comparison for individual or combined groups of tests, reporting balanced accuracy and area under the ROC curve (Bal.Acc./AUC)

	Controls vs. AD	Controls vs. FTD	Controls vs. VAD	Controls vs. DLB	AD vs. FTD	AD vs. VaD	AD vs. DLB	FTD vs. VAD	FTD vs. DLB	VaD vs. DLB
Cognitive tests	95/98	87/94	95/98	92/97	64/70	62/69	68/76	59/64	62/68	57/58
CSF	83/91	57/67	60/63	59/65	85/92	76/87	76/83	60/64	60/65	46/48
MRI	86/93	86/92	87/94	79/84	70/79	80/90	65/75	82/89	75/84	78/84
Cognitive tests + CSF	96/99	86/94	95/98	92/96	82/86	78/81	78/85	66/67	64/71	53/56
Cognitive tests + MRI	96/99	92/97	96/98	92/97	73/81	78/87	75/80	80/89	77/85	77/82
CSF + MRI	89/95	87/91	87/92	79/83	83/89	83/91	79/83	83/89	76/85	76/83
Cognitive tests + CSF + MRI	97/99	91/97	96/98	92/97	85/90	83/90	79/87	83/89	78/86	77/81

Abbreviations: ROC, receiver operating characteristic; CSF, cerebrospinal fluid; MRI, magnetic resonance imaging; AD, Alzheimer's disease; FTD, frontotemporal dementia; VaD, vascular dementia; DLB, dementia with Lewy bodies; AUC, area under the ROC curve; Bal.Acc., balanced accuracy.

NOTE. The table shows for each pairwise comparison of two diagnostic groups the balanced accuracy and AUC (Bal.Acc./AUC) achieved when combining all diagnostic tests for each individual group of tests: cognitive tests, CSF, and MRI. Moreover, the balanced accuracy and AUC when combining the groups of tests are reported. Balanced accuracies 85%–100% are highlighted in dark green. The gradually lighter shades of green indicate lower balanced accuracy with white being at or below 50. Both balanced accuracy and AUC are reported as percentage values (%).

Table 4

The optimal sets of diagnostic tests for each pairwise comparison

Pairwise comparison	Cognitive tests	CSF	MRI	Bal.Acc./AUC
Controls vs. AD	MMSE, Memory (recall)		ROI-based grading	95/99
Controls vs. FTD	MMSE, Animal fluency		Global VBM	92/97
Controls vs. VaD	Animal fluency		Vascular burden, temporal cortex	95/99
Controls vs. DLB	TMT-B, Memory (learning)		ROI-based grading, anterior versus posterior index	94/98
AD vs. FTD		A β ₄₂ , p-tau		88/93
AD vs. VaD	TMT-A	p-tau	Vascular burden, medial temporal cortex, ROI-based grading	86/94
AD vs. DLB	Memory (recall), TMT-A	A β ₄₂ , total tau	ROI-based grading, temporal cortex	82/88
FTD vs. VaD			Vascular burden, global TBM index	89/96
FTD vs. DLB	TMT-A	A β ₄₂	Anterior versus posterior index	84/90
VaD vs. DLB			Vascular burden	85/92

Abbreviations: AD, Alzheimer's disease; FTD, frontotemporal dementia; VaD, vascular dementia; DLB, dementia with Lewy bodies; MMSE, Mini-Mental State Examination; TMT, Trail Making Test; A β ₄₂, β amyloid 1–42; p-tau, tau phosphorylated at threonine 181; MRI, magnetic resonance imaging; VBM, voxel-based morphometry; TBM, tensor-based morphometry; CSF, cerebrospinal fluid; WMH, white matter hyperintensity; AUC, area under the receiver operating characteristic curve; Bal.Acc., balanced accuracy.

NOTE. The table shows the combinations of diagnostic tests that achieve the highest AUC in pairwise comparison of diagnostic groups.

NOTE. MRI: volumes are defined from image segmentations produced by a multi-atlas segmentation algorithm. ROI-based grading: based on hippocampus and frontal lobe regions ROI. Vascular burden: based on three features—volume of WMH, volume of cortical infarcts, and volume of lacunar infarcts. Anterior versus posterior index: index between anterior and posterior weighted volumes.

using the classifier to provide an overview of the individual and optimized combined performance of cognitive tests, CSF biomarkers, and automated MRI features with regard to discriminating the most common types of dementias in a larger multicenter data set.

Looking at the individual diagnostic tests, we found, as expected, that cognitive tests were most important for syndrome diagnosis, separating controls from dementia. Cognitive tests had little value in differentiating between the different dementia types, probably due to considerable overlap in clinical presentation within the diagnostic groups [8,38,39]. Furthermore, we found CSF biomarkers highly useful in separating AD from controls and other dementia types, but of little value when differentiating between VaD, DLB, and FTD, which is in line with observational studies and previous studies using the DSI classifier [37,40,41]. The higher accuracy of CSF biomarkers in the separation of AD from FTD, as compared with AD from DLB or VaD, is probably due to more frequent mixed neuropathology in the latter diagnoses [42]. Finally, most of the automated MRI features performed with high accuracy when separating controls and dementias. By definition, vascular burden was the best MRI feature to define VaD from all other diagnostic groups [27]. Furthermore, the anterior versus posterior index had a high accuracy for identifying FTD from AD and DLB. Moreover, global VBM and TBM indices and ROI-based grading features performed with some of the highest accuracies for DLB and FTD differentiations, in line with previous study [43]. These advanced imaging biomarkers are able to focus on the most relevant areas and local patterns extending the capacity of conventional volumetric features. They, however, also provide huge amounts of data. Previous studies have demonstrated that the highest increase in the diagnostic performance of MRI was achieved by selecting specific automated MRI features [27,44–46].

When examining the combination of tests for pairwise differential diagnosis of dementias, we found that, especially, adding automated MRI features to either cognitive tests or CSF biomarkers increased the accuracy [47]. Moreover, we explored the optimal combination of diagnostic tests, which seem to improve the accuracies further, although performance results are slightly overestimated due to not applying cross-validation to this final analysis. Overall, only a few of the cognitive tests increased the accuracy and were selected for the optimal combinations between the diagnostic dementia groups. In line with the literature, combinations for separation of AD from other types of dementia showed dominance of CSF biomarkers [47]. Except for one pairwise comparison, automated MRI features were included in all optimized combinations of diagnostic tests. Predominantly, the selected MRI features were the more advanced features, such as VBM, TBM, vascular burden, anterior versus posterior index, and ROI-based grading features, rather than standard volumetric measures. The optimized combinations of diagnostic tests for comparison of AD vs. VaD and AD vs. DLB contained a higher number of diagnostic tests, which could be due to frequent overlap of neuropathology or the fact that few specific biomarkers are available for DLB [48]. Finally, we found all optimized diagnostic sets to be different stressing the importance and complexity of weighing and combining biomarkers and tests correctly in clinical practice.

The strengths of our study are the large multicenter cohort, containing patients diagnosed with the most frequent types of dementia. Furthermore, a standardized, thorough clinical workup of all patients in each center, and the fact that using multicenter data, adds to the generalizability of our results. We used data that were typical of memory clinics not optimized or imputed but varied and incomplete. Because we aimed to study biomarkers used in daily

practice, it is essential the analysis method can deal with missing data or data of suboptimal quality [37,40,49].

One limitation of this study was that a comparable independent validation cohort was not available. Using cross-validation may lead to overoptimistic accuracy estimates, but choosing either the Amsterdam Dementia Cohort or PredictND cohort as an independent cohort would have compromised the precision of accuracy estimates due to the relatively low number of patients with FTD, DLB, and VaD. Likewise, an independent cohort is also needed for validating the optimal combination of diagnostic tests defined by feature selection. However, the main purpose of the study was not to provide precise estimates about the diagnostic accuracy for different diagnostic comparisons but rather to present clinically useful information about how different diagnostic tests, alone or combined, contribute to differential diagnostics of dementia. Additional diagnostic tests, such as amyloid- and 18F-fluorodeoxyglucose–positron emission tomography, were not included in the study but could have provided even more insight into the performance of diagnostic tests. Another limitation is that generally 20%–40% of patients with dementia have more than one underlying pathologies that are accounted for neither in the clinical diagnosis nor the classifier [48]. This could presumably have underestimated the accuracy. The fact that clinical diagnosis was used as reference diagnosis is also associated with some constraints as agreement between clinical diagnosis and postmortem neuropathological diagnosis has been reported to be 70%–90% in dementias [50,51]. A confirmed neuropathological diagnosis would therefore have been preferable. Furthermore, cognitive tests and CSF biomarkers were used to determine the clinical diagnosis. However, the automated features were computed for this study and had not been used for clinical diagnosis. Finally, although FTD is a heterogenic disease including subtypes, behavioral variant FTD, semantic dementia, and progressive nonfluent aphasia were evaluated as one diagnostic group in this study. With a larger group of patients with FTD, it would be interesting to explore the performance of the diagnostic tests for each of these subtypes separately.

This study has important clinical implications; evidence-based medicine is struggling with the challenge of comparing multiple sources of data in differential diagnosis of dementia. Our study demonstrates how combinations of diagnostic tests for pairwise comparison of diagnostic groups can be studied using clinical decision support system, though, still based on group-level information. However, applying classifiers like this to the diagnostic assessment in clinical practice could assist clinicians in the simultaneous evaluation of a wide range of diagnostic tests while evaluating different types of dementia on a personalized level [19].

In conclusion, based on a data-driven approach this research shows that different diagnostic tests have their distinct roles in differential diagnostics of dementias. Our re-

sults also indicate that combining different tests seems to increase accuracy in differentiating several diagnostic groups. Together with knowledge from conventional studies, these results could help clinicians to prioritize biomarkers and improve clinical practice for patients with dementia.

Acknowledgments

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreements no 611005 (PredictND). For development of the PredictND tool, the VTT Technical Research Center of Finland Ltd has received funding from European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreements 601055 (VPH-DARE@IT), 224328, and 611005. The PredictND consortium consisted of collaborates from the VTT Technical Research Center of Finland, GE Healthcare Ltd, Imperial College London, Alzheimer Europe, Alzheimer Center–VU University Medical Center, Amsterdam, the Netherlands, the Danish Dementia Research Center, Copenhagen University Hospital, Denmark, the Department of Gerontology and Geriatrics of the University of Perugia, "S. Maria della Misericordia" Hospital of Perugia, Italy, the Department of Neurology from the University of Eastern Finland.

Author disclosures: M. Bruun, H.F.M.R.-M., M. Baroni, L.G., A.W.L., A.M.R., T.U., A.T., D.R., M.v.G., K.S.R., G.W., P.M., and S.G.H. report no disclosures. F.B. is supported by the NIHR UCLH Biomedical Research Centre. H.S. has served in advisory boards for AC Immune, MSD, and Orion Pharma. P.S. has served as a consultant for Wyeth-Elan, Genentech, Danone, and Novartis and received funding for travel from Pfizer, Elan, Janssen, and Danone Research. J.L. and J.K. are shareholders in Combinostics Oy that owns the following IPR related to the patent: (1) J. Koikkalainen and J. Lotjonen—A method for inferring the state of a system, US7,840,510 B2, PCT/FI2007/050277. (2) J. Lotjonen, J. Koikkalainen, and J. Mattila—State Inference in a heterogeneous system, PCT/FI2010/050545, FI20125177. W.M.v.d.F. performs contract research for Biogen. Research programs of W.M.v.d.F. have been funded by ZonMW, NWO, EU-FP7, Alzheimer Nederland, CardioVascular Onderzoek Nederland, Stichting Dioraphte, Gieskes-Strijbis fonds, Boehringer Ingelheim, Piramal Neuroimaging, Roche BV, Janssen Stellar, and Combinostics. All funding is paid to her institution.

Authors' contributions: M. Bruun drafted the manuscript and analyzed/interpreted data. H.F.M.R.-M. drafted the manuscript and analyzed/interpreted data. J.K. revised the manuscript and analyzed/interpreted data. M. Baroni, L.G., A.W.L., F.B., A.M.R., T.U., A.T., D.R., M.v.G., K.S.F., G.W., P.M., and P.S. revised the manuscript and interpreted data. H.S. contributed to data acquisition and management and revision of the manuscript. J.L. revised the manuscript

and analyzed/interpreted data. S.G.H. drafted the manuscript, analyzed/interpreted data, and supervised the project. W.M.v.d.F. drafted the manuscript, analyzed/interpreted data, and supervised the project.

Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.dadm.2018.07.003>.

RESEARCH IN CONTEXT

1. Systematic review: We searched the literature using scientific databases focusing on biomarkers for differential dementia diagnostics and classifiers. Most studies focused on either two types of dementia or few types of diagnostic tests. There is a paucity of knowledge on combining biomarkers optimally in differential diagnosis of dementia in clinical practice.
2. Interpretation: Using a data-driven classifier, we demonstrated accuracies for individual performance and optimal combinations of cognitive tests, cerebrospinal fluid biomarkers, and automated magnetic resonance imaging-derived features for pairwise differential diagnosis of dementias, focusing on Alzheimer's disease, frontotemporal dementia, dementia with Lewy bodies, and vascular dementia. Through new machine learning-based technologies, our research provides a more holistic view of the combinatorial performance of diagnostic tests in clinical differential diagnostic dilemmas.
3. Future directions: This research shows that the optimal combinations of diagnostic tests differ for each pairwise separation of diagnostic groups. Future research should focus on confirming these optimal combinations and translate this knowledge into clinical practice.

- [3] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 2011;7:263–9.
- [4] Roman GC, Tatemonchi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JH, et al. Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology* 1993;43:250–60.
- [5] Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 2011;134:2456–77.
- [6] McKeith IG, Boeve BF, Dickson DW, Halliday G, Taylor JP, Weintraub D, et al. Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology* 2017;89:88–100.
- [7] Smits LL, van Harten AC, Pijnenburg YA, Koedam EL, Bouwman FH, Sisternans N, et al. Trajectories of cognitive decline in different types of dementia. *Psychol Med* 2015;45:1051–9.
- [8] Burrell JR, Piguet O. Lifting the veil: How to use clinical neuropsychology to assess dementia. *J Neurol Neurosurg Psychiatr* 2015; 86:1216–24.
- [9] Schoonenboom NS, Reesink FE, Verwey NA, Kester MI, Teunissen CE, van de Ven PM, et al. Cerebrospinal fluid markers for differential dementia diagnosis in a large memory clinic cohort. *Neurology* 2012;78:47–54.
- [10] Blennow K, Dubois B, Fagan AM, Lewczuk P, De Leon MJ, Hampel H. Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease. *Alzheimers Dement* 2015; 11:58–69.
- [11] Burton EJ, Barber R, Mukaetova-Ladinska EB, Robson J, Perry RH, Jaros E, et al. Medial temporal lobe atrophy on MRI differentiates Alzheimer's disease from dementia with Lewy bodies and vascular cognitive impairment: A prospective study with pathological verification of diagnosis. *Brain* 2009;132:195–203.
- [12] Koedam EL, Lehmann M, van der Flier WM, Scheltens P, Pijnenburg YA, Fox N, et al. Visual assessment of posterior atrophy development of a MRI rating scale. *Eur Radiol* 2011;21:2618–25.
- [13] Rhodius-Meester HFM, Benedictus MR, Wattjes MP, Barkhof F, Scheltens P, Muller M, et al. MRI visual ratings of brain atrophy and white matter hyperintensities across the spectrum of cognitive decline are differently affected by age and diagnosis. *Front Aging Neurosci* 2017;9:117.
- [14] Westman E, Muehlboeck JS, Simmons A. Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *Neuroimage* 2012; 62:229–38.
- [15] Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 2011;55:856–67.
- [16] Mattila J, Koikkalainen J, Virkki A, van GM, Lotjonen J. Design and application of a generic clinical decision support system for multiscale data. *IEEE Trans Biomed Eng* 2012;59:234–40.
- [17] van der Flier WM, Pijnenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: The Amsterdam Dementia Cohort. *J Alzheimers Dis* 2014; 41:313–27.
- [18] van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing research to optimize care. *J Alzheimers Dis* 2018; 62:1091–111.
- [19] Bruun M, Gjerum L, Frederiksen KS, Rhodius-Meester HFM, Baroni M, Lemstra E, et al. Data-driven diagnosis of dementia disorders: The predictND validation study. *Alzheimers Dement* 2017; 13:405–7.
- [20] Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. *Neurology* 2011;76:1006–14.

References

- [1] World Health Organization. Global action plan on the public health response to dementia 2017–2025. World Health Organization 2017:1–44. Available at: http://www.who.int/mental_health/neurology/dementia/action_plan_2017_2025/en.
- [2] Lobo A, Launer LJ, Fratiglioni L, Andersen K, Di Carlo A, Breteler MM, et al. Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. *Neurologic Diseases in the Elderly Research Group*. *Neurology* 2000;54:S4–9.

- [21] McKeith IG, Dickson DW, Lowe J, Emre M, O'Brien JT, Feldman H, et al. Diagnosis and management of dementia with Lewy bodies: Third report of the DLB Consortium. *Neurology* 2005;65:1863–72.
- [22] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975;12:189–98.
- [23] Schmidt M. Rey auditory verbal learning test: A handbook. Los Angeles, CA: Western Psychological Services; 1996.
- [24] Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989;39:1159–65.
- [25] Reitan R. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills* 1958;8:271–6.
- [26] Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. *J Int Neuropsychol Soc* 2006;12:80–9.
- [27] Koikkalainen J, Rhodius-Meester H, Tolonen A, Barkhof F, Tijms B, Lemstra AW, et al. Differential diagnosis of neurodegenerative diseases using structural MRI data. *NeuroImage Clin* 2016;11:435–49.
- [28] Lotjonen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* 2010;49:2352–65.
- [29] Bruun M, Koikkalainen J, Rhodius-Meester HFM, Baroni M, Gjerum L, van Gilsf M, et al. Detecting frontotemporal dementia using a novel MRI imaging biomarker: The Anterior vs. Posterior Index. Paper presented at: Alzheimer's Association International Conference 2018; July 22–26, 2018; Chicago, IL, USA.
- [30] Ashburner J, Friston KJ. Voxel-based morphometry—the methods. *Neuroimage* 2000;11:805–21.
- [31] Koikkalainen J, Lotjonen J, Thurfjell L, Rueckert D, Waldemar G, Soininen H. Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease. *Neuroimage* 2011;56:1134–44.
- [32] Tong T, Wolz R, Coupe P, Hajnal JV, Rueckert D. Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. *Neuroimage* 2013;76:11–23.
- [33] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van GM, Waldemar G, et al. A disease state fingerprint for evaluation of Alzheimer's disease. *J Alzheimers Dis* 2011;27:163–76.
- [34] Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume. *Neuroimage* 2004;23:724–38.
- [35] Cole TJ, Green PJ. Smoothing reference centile curves: The LMS method and penalized likelihood. *Stat Med* 1992;11:1305–19.
- [36] Cluitmans L, Mattila J, Runtu H, van Gils M, Lotjonen J. A MATLAB toolbox for classification and visualization of heterogenous multi-scale human data using the Disease State Fingerprint method. *Stud Health Technol Inform* 2013;189:77–82.
- [37] Tolonen A, Rhodius-meester HFM, Bruun M, Koikkalainen J, Barkhof F, Lemstra AW, et al. Data-Driven Differential Diagnosis of Dementia Using Multiclass Disease State Index Classifier. *Front Aging Neurosci* 2018;10:1–11. <https://doi.org/10.3389/fnagi.2018.00111>.
- [38] Sitek EJ, Barczak A, Harciarek M. Neuropsychological assessment and differential diagnosis in young-onset dementias. *Psychiatr Clin North Am* 2015;38:265–79.
- [39] Scheltens NME, Tijms BM, Koene T, Barkhof F, Teunissen CE, Wolfgruber S, et al. Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts. *Alzheimers Dement* 2017;13:1226–36.
- [40] Rhodius-Meester HF, Koikkalainen J, Mattila J, Teunissen CE, Barkhof F, Lemstra AW, et al. Integrating biomarkers for underlying Alzheimer's disease in mild cognitive impairment in daily practice: Comparison of a clinical decision support system with individual biomarkers. *J Alzheimers Dis* 2015;50:261–70.
- [41] Munoz-Ruiz MA, Hartikainen P, Hall A, Mattila J, Koikkalainen J, Herukka SK, et al. Disease state fingerprint in frontotemporal degeneration with reference to Alzheimer's disease and mild cognitive impairment. *J Alzheimers Dis* 2013;35:727–39.
- [42] Toledo JB, Brettschneider J, Grossman M, Arnold SE, Hu WT, Xie SX, et al. CSF biomarkers cutoffs: The importance of coincident neuropathological diseases. *Acta Neuropathol* 2012;124:23–35.
- [43] Munoz-Ruiz MA, Hartikainen P, Koikkalainen J, Wolz R, Julkunen V, Niskanen E, et al. Structural MRI in frontotemporal dementia: Comparisons between hippocampal volumetry, tensor-based morphometry and voxel-based morphometry. *PLoS One* 2012;7:e52531.
- [44] Moller C, Hafkemeijer A, Pijnenburg YA, Rombouts SA, van der Grond J, Dopper E, et al. Joint assessment of white matter integrity, cortical and subcortical atrophy to distinguish AD from behavioral variant FTD: A two-center study. *NeuroImage Clin* 2015;9:418–29.
- [45] Canu E, Agosta F, Mandic-Stojmenovic G, Stojkovic T, Stefanova E, Inuggi A, et al. Multiparametric MRI to distinguish early onset Alzheimer's disease and behavioural variant of frontotemporal dementia. *NeuroImage Clin* 2017;15:428–38.
- [46] Wolz R, Julkunen V, Koikkalainen J, Niskanen E, Zhang DP, Rueckert D, et al. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease. *PLoS One* 2011;6:e25446.
- [47] Schoonenboom NS, van der Flier WM, Blankenstein MA, Bouwman FH, Van Kamp GJ, Barkhof F, et al. CSF and MRI markers independently contribute to the diagnosis of Alzheimer's disease. *Neurobiol Aging* 2008;29:669–75.
- [48] Zekry D, Hauw JJ, Gold G. Mixed dementia: Epidemiology, diagnosis, and treatment. *J Am Geriatr Soc* 2002;50:1431–8.
- [49] Hall A, Munoz-Ruiz M, Mattila J, Koikkalainen J, Tsolaki M, Mecocci P, et al. Generalizability of the disease state index prediction model for identifying patients progressing from mild cognitive impairment to Alzheimer's disease. *J Alzheimers Dis* 2015;44:79–92.
- [50] Jellinger KA. Accuracy of clinical criteria for AD in the Honolulu-Asia Aging Study, a population-based study. *Neurology* 2002;58:989–90.
- [51] Kazee AM, Eskin TA, Lapham LW, Gabriel KR, McDaniel KD, Hamill RW. Clinicopathologic correlates in Alzheimer disease: Assessment of clinical and pathologic diagnostic criteria. *Alzheimer Dis Assoc Disord* 1993;7:152–64.