

Combination of Structural MRI and FDG-PET of the Brain Improves Diagnostic Accuracy in Newly Manifested Cognitive Impairment in Geriatric Inpatients

Kerstin Ritter^{a,*}, Catharina Lange^b, Martin Weygandt^a, Anja Mäurer^c, Anna Roberts^c, Melanie Estrella^d, Per Suppa^{b,e}, Lothar Spies^c, Vikas Prasad^b, Ingo Steffen^b, Ivayla Apostolova^f, Daniel Bittner^g, Mehmet Gövercin^d, Winfried Brenner^b, Christine Mende^h, Oliver Petersⁱ, Joachim Seybold^{c,j}, Jochen B. Fiebach^k, Elisabeth Steinhagen-Thiessen^l, Harald Hampel^m, John-Dylan Haynes^a and Ralph Buchert^{b,*}

^a*Berlin Center for Advanced Neuroimaging, Bernstein Center for Computational Neuroscience, Charité - Universitätsmedizin Berlin, Berlin, Germany*

^b*Department of Nuclear Medicine, Charité - Universitätsmedizin Berlin, Berlin, Germany*

^c*Evangelisches Geriatriezentrum Berlin, Berlin, Germany*

^d*Geriatric Research Group, Department of Geriatric Medicine, Charité-Universitätsmedizin Berlin, Berlin, Germany*

^e*Jung Diagnostics GmbH, Hamburg, Germany*

^f*Department of Radiology and Nuclear Medicine, University Hospital Magdeburg, Magdeburg, Germany*

^g*Department of Neurology, University Hospital Magdeburg, Magdeburg, Germany*

^h*Vivantes Ida Wolff Krankenhaus, Berlin, Germany*

ⁱ*Department of Psychiatry and Psychotherapy Charité Benjamin Franklin, Charité - Universitätsmedizin Berlin, Berlin, Germany*

^j*Department of Internal Medicine/Infectious Diseases and Pulmonary Medicine, Charité - Universitätsmedizin Berlin, Germany*

^k*Center for Stroke Research Berlin, Charité, Berlin, Germany*

^l*Lipid Clinic at the Interdisciplinary Metabolism Center, Charité - Universitätsmedizin, Berlin, Germany*

^m*AXA Research Fund & UPMC Chair, Sorbonne Universités, Université Pierre et Marie Curie, Paris 06, Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A) & Institut du Cerveau et de la Moelle épinière (ICM), Département de Neurologie, Hôpital de la Pitié-Salpêtrière, Paris, France*

Accepted 4 July 2016

Abstract.

Background: The cause of cognitive impairment in acutely hospitalized geriatric patients is often unclear. The diagnostic process is challenging but important in order to treat potentially life-threatening etiologies or identify underlying neurodegenerative disease.

Objective: To evaluate the add-on diagnostic value of structural and metabolic neuroimaging in newly manifested cognitive impairment in elderly geriatric inpatients.

*Correspondence to: Kerstin Ritter, PhD and Ralph Buchert, PhD, Charitéplatz 1, 10117 Berlin, Germany. Tel.: +49 30 450 539364; Fax: +49 30 2093 6758; E-mail: kerstin.ritter@bccn-

berlin.de (K. Ritter); Tel.: +49 30 450 627059; Fax: +49 30 450 7527959; E-mail: ralph.buchert@charite.de (R. Buchert).

Methods: Eighty-one inpatients (55 females, 81.6 ± 5.5 y) without history of cognitive complaints prior to hospitalization were recruited in 10 acute geriatrics clinics. Primary inclusion criterion was a clinical hypothesis of Alzheimer's disease (AD), cerebrovascular disease (CVD), or mixed AD+CVD etiology (MD), which remained uncertain after standard diagnostic workup. Additional procedures performed after enrollment included detailed neuropsychological testing and structural MRI and FDG-PET of the brain. An interdisciplinary expert team established the most probable etiologic diagnosis (non-neurodegenerative, AD, CVD, or MD) integrating all available data. Automatic multimodal classification based on Random Undersampling Boosting was used for rater-independent assessment of the complementary contribution of the additional diagnostic procedures to the etiologic diagnosis.

Results: Automatic 4-class classification based on all diagnostic routine standard procedures combined reproduced the etiologic expert diagnosis in 31% of the patients ($p=0.100$, chance level 25%). Highest accuracy by a single modality was achieved by MRI or FDG-PET (both 45%, $p \leq 0.001$). Integration of all modalities resulted in 76% accuracy ($p \leq 0.001$).

Conclusion: These results indicate substantial improvement of diagnostic accuracy in uncertain de novo cognitive impairment in acutely hospitalized geriatric patients with the integration of structural MRI and brain FDG-PET into the diagnostic process.

Keywords: Cognitive impairment, geriatric inpatients, magnetic resonance imaging, multimodal classification, positron emission tomography

INTRODUCTION

Dementia disorders are highly prevalent in the elderly and on the rise due to worldwide population aging. The clinical diagnosis of a late prodromal or dementia stage neurodegenerative disease including Alzheimer's disease (AD) is a very serious event for patients and their relatives [1]. Individuals aged 60 or older are most afraid of being affected by AD, even more than of cancer and stroke combined [2]. Therefore, it is important to control the rate of false positive AD diagnoses [3]. In standard clinical routine, the diagnostic process relies on descriptive symptom-based criteria [4]. Diagnosis based on symptoms alone, however, is prone to be inaccurate, particularly at earlier clinical stages when symptoms are emerging, subtle or mild, and in cases with inconclusive, incomplete, and/or atypical clinical presentation [5].

According to newly updated criteria, accuracy of the etiologic diagnosis of clinically uncertain cognitive impairment (CUCI) can be improved by complementing symptom-based criteria with evidence of the underlying pathophysiology or of characteristic structural, functional, and metabolic topographical alterations based on biomarkers derived from magnetic resonance imaging (MRI), positron emission tomography (PET), or cerebrospinal fluid (CSF) assays [6, 7]. The vast majority of studies on the use of these biomarkers for the diagnosis of AD, however, have focused on memory clinic outpatient settings and employed restrictive eligibility criteria resulting in rather highly selected patient samples that might not be representative of the 'typical' patient in the community [8]. Trial participants recruited in urban major academic centers tend to be

highly educated and to be in better than average physical condition [8]. The role of biomarkers in more 'difficult' and primary care settings is still poorly investigated. For example, initial clinical experience with the combination of different biomarkers indicates that the fraction of patients with completely congruent biomarkers might be surprisingly small in routine patient care [9].

A particularly difficult clinical setting is CUCI in patients that are admitted and hospitalized in a geriatric hospital or unit for an acute or subacute illness. Compared to the typical outpatient of a psychiatric or neurological memory clinic, acutely hospitalized geriatric patients suffer from more complex medical conditions and/or comorbidities, often associated with considerable pain. In addition, they are on more extensive medication, often including centrally acting drugs with considerable tolerability problems and side effects impacting cognition. Typical presentations include patients with prolonged cognitive impairment during recovery from surgical treatment on a geriatric unit, although they had no history of cognitive problems prior to admission and hospitalization. After ruling out delirium and depressive episode, the suspicion of a first clinical manifestation of a neurodegenerative disease should be assessed before the patient is finally discharged from the acute care geriatrics unit [10].

We postulate that biomarkers are particularly useful in these difficult cases. The present study evaluated the added value of structural MRI [11] and PET with 2-[F-18]-fluoro-2-deoxy-D-glucose (FDG) [12] together with additional neuropsychological testing over standard diagnostic workup in newly manifested, clinically uncertain cognitive impairment in

elderly patients hospitalized in a geriatrics unit due to an acute or subacute indication.

MATERIALS AND METHODS

Patients

All data were derived from the prospective clinical trial ‘Comparison and integration of modalities in the early and differential diagnosis of dementing disorders in hospitalized geriatric patients: a prediction study’ (acronym: iDSS (integrated Decision Support System)).

Eligibility criteria that led to the inclusion of 109 patients are given in an appendix. Six of the 109 patients dropped-out after inclusion but prior to the first study-related examination. Structural MRI was successfully completed in 90 patients, FDG-PET in 97. We included all patients ($n=81$) with (i) both structural MRI and FDG-PET (no patients were excluded due to limited image quality) and (ii) a etiological diagnosis (as described below) of either a non-neurodegenerative etiology, cerebrovascular disease (CVD), AD, or mixed CVD+AD etiology (MD). Patients with a baseline diagnosis of a neurodegenerative disease other than AD were excluded and, therefore, did not affect the outcomes of this study. The study flowchart is shown in Fig. 1.

Standard diagnostic workup

Standard diagnostic workup included patient history (including education, lifestyle, and environmental factors), physical/neurological examination, standard neuropsychological testing, standard blood/urine laboratory tests (including insulin, different vitamins, and inflammation markers), and ApoE genotyping. Standard neuropsychological testing consisted of: the Mini-Mental State Examination (MMSE) [13], the DemTect psychometric screening tool [14], Clinical Dementia Rating (CDR) [15], the behavioral pathology in Alzheimer’s disease rating scale (BEHAVE-AD) [16], and a short version of the geriatric depression scale for inpatients (GDS-K) [17]. Standard diagnostic workup did not include additional dementia-specific tests. Structural brain imaging (CT or MRI) performed prior to inclusion was available only in very few patients where it had been performed to exclude treatable causes of the cognitive impairment such as subdural hematoma, brain tumor or normal pressure hydrocephalus.

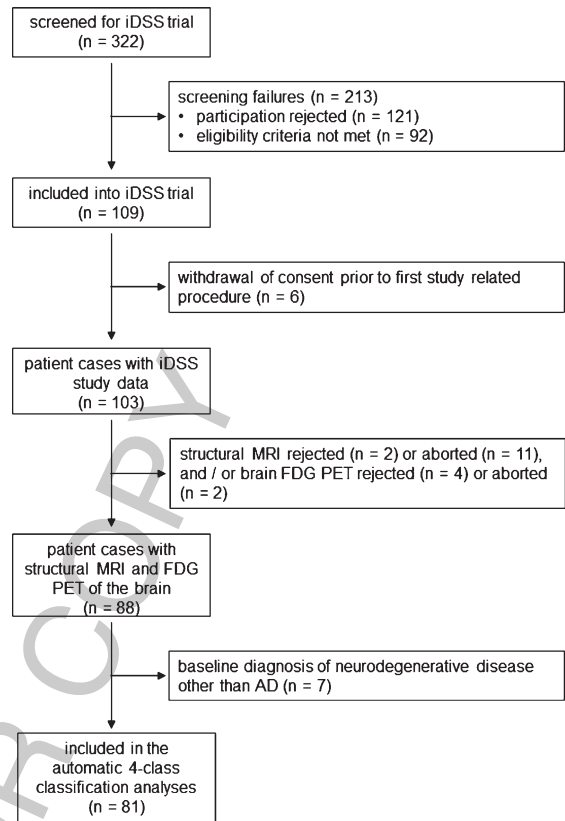


Fig. 1. Study flowchart.

Additional diagnostic procedures

Additional diagnostic procedures performed as part of the clinical study included further neuropsychological testing, structural MRI and FDG-PET.

Additional neuropsychological testing comprised the German version of the test battery of the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) including trail making test A+B and phonematic fluency (CERAD-plus w/o MMSE) [18], a multiple choice vocabulary test to estimate premorbid intelligence (MWT-A) [19], the Montgomery-Asberg Depression Rating Scale (MADRS) [20], and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) [21].

MRI of the brain was performed with the same 3 Tesla MR Scanner (Siemens Trio) in all patients. The structural sequences included 3-dimensional T1-weighted MPRAGE ($1 \times 1 \times 1 \text{ mm}^3$), T2-weighted FLAIR (in-plane 1.2 mm, slice thickness 2.5 mm), a T2*-weighted sequence (in-plane 0.7 mm, slice thickness 5 mm), and ToF non-contrast enhanced MRA. Structural MRI was interpreted visually

on a computer monitor of a PACS workstation. Visual interpretation of mesio-temporal atrophy was supported by quantitative analysis of regional grey matter volume combining voxel-based morphometry and atlas-based hippocampal volumetry (Biometrica AD[®], jung diagnostics GmbH, Hamburg, Germany). Visual interpretation of MRI for vascular disease assessed large vessel strokes, recent small subcortical infarcts, chronic lacunes, cerebral microbleeds/siderosis, and white matter hyperintensities (Supplementary Material 1).

A PET scan of the brain of 15 min duration was acquired 40 ± 5 min after i.v. administration of about 200 MBq FDG according to common guidelines [22]. The same PET/CT scanner (Philips Gemini TF 16) was used in all patients. PET emission data were reconstructed into 15 frames each of 1 min duration using the iterative reconstruction algorithm of the PET/CT scanner software. Spatial resolution in the reconstructed images was about 7 mm full-width-of-half-maximum. Image processing including post acquisition inter-frame motion correction and voxel-based statistical testing was performed using the Statistical Parametric Mapping software (version SPM8) as described by Lange and colleagues [23]. In short, the 'realign' tool of SPM8 with the first frame as reference was used for inter-frame motion correction. The magnitude of the motion from frame to frame was estimated by tracking 5 predefined reference points in the brain. Frames with a motion amplitude >4 mm (about half the spatial resolution) were discarded in order to avoid relevant errors by mismatch between PET and low-dose CT for attenuation correction [24]. A motion-corrected static uptake image was obtained by summing the remaining frames after realignment.

A sample report used for visual interpretation of FDG-PET is provided in Supplementary Material 2.

Participation in lumbar puncture for analysis of CSF was optional. Only 10 of the 81 patients consented in lumbar puncture so that CSF was not included in the analyses.

Etiologic diagnosis by experts

Each participant was classified into one of the following five etiology classes: non-neurodegenerative etiology, AD, CVD, MD, and neurodegenerative etiology different from AD. Classification consensus was based on all available data, i.e., from both standard diagnostic workup and additional diagnostic procedures. The consensus panel consisted of

an inter-disciplinary team of academic experts for internal medicine (A.R.), neuropsychology (A.M.), nuclear medicine (R.B.), and neuroradiology (J.F.), all experienced in the diagnosis of neurodegenerative diseases. Diagnostic criteria for AD [7] and vascular dementia [25] were taken into account as far as possible (s. Discussion). A patient was classified as MD if there was evidence for both AD and CVD as cause of the CUCI. The 'non-neurodegenerative etiology' class included reduced general health, depression, prolonged effect of delirium and hippocampal sclerosis of aging [26, 27]. Patients in the 'non-AD neurodegeneration' class were excluded from the analyses (Fig. 1).

Automatic etiologic classification

In addition to the etiological diagnosis by the experts, etiological classification was also performed fully automatically using a machine learning method that accounts for complex relationships between data [28]: RUSBoost ('Random Undersampling Boosting'), an optimized version of the AdaBoost ('Adaptive Boosting') algorithm [29]. In AdaBoost weak learners are iteratively adapted so that training examples which have been misclassified in the previous iteration become a stronger weight in the next iteration [30]. The final classification is then done by a weighted vote of all built models. As weak learners we chose decision trees, which is a common choice for AdaBoost. Random undersampling is introduced into AdaBoost in order to account for unbalanced class size [29]. The Matlab2013 implementation of RUSBoost was used with default parameters. The number of trees grown per training cycle was set to 100.

Classification analyses were performed separately for (i) each single feature (e.g., MMSE sum score), (ii) each single modality (e.g., PET), (iii) all standard features combined, (iv) all features from both standard and additional modalities combined, and (v) a set of expert features. For each of these analyses, overall accuracy of the full 4-class classification was determined. In addition, accuracy for detection of each of the single etiological classes (versus all other classes) was obtained. In order to assess the performance in detection of neurodegeneration, we evaluated the accuracy of automatic discrimination between the patients with pure AD and the patients with MD combined into one group versus all others (patients with pure CVD and patients with non-neurodegenerative etiology). In order to assess the

performance in detection of cerebrovascular disease, we evaluated the accuracy of automatic discrimination between the patients with pure CVD and the patients with MD combined into one group versus all others (patients with pure AD and patients with non-neurodegenerative etiology).

To estimate the generalization error for all classification analyses, we used 5-fold cross-validation, which means that 4/5 of the data were used for training and 1/5 for testing to ensure complete independence between training and test data. This was repeated such that each 1/5 of the data was used once as the testing data [30]. Using this procedure, a class label was generated for each patient, which then was compared to the expert consensus diagnosis. The cross-validation procedure was repeated 30 times to produce stable results.

To test for statistical significance, a permutation test was used, in which the labels were randomly permuted for 1000 times and then the number of cases was counted in which the accuracy of permuted labels exceeded the true accuracy. The p -value was then determined as the fraction of these cases relative to the total number of permutations. The lowest achievable p -value was 0.001.

Features for automatic classification

Features for automatic etiologic classification were categorized into ‘standard features’ and ‘additional features’. The standard features included demographic data (‘demo’: age, gender and education, 3 features), clinical data (‘clinical’, 32 features), laboratory data (‘lab’, 22 features), data from the standard neuropsychological tests (‘standardNP’: DemTect, GDS, CDR, BEHAVE and MMSE sum score, 5 features), and ApoE genotype parametrized as the number of ApoE4 alleles (‘ApoE’, 1 feature). The additional features included data from additional neuropsychological testing (‘additionalNP’, 14 features), MRI features separated into ‘vascularMRI’ features comprising visual scores for CVD (large vessel stroke, recent small subcortical infarct, chronic lacunes, cerebral microbleeds/siderosis, white matter hyperintensities [25], 24 features) and ‘volumetricMRI’ features comprising grey matter volume of various predefined brain regions [31, 32] (14 features), and mean FDG uptake in predefined brain regions [23] (‘PET’, 26 features). Regional FDG uptake was scaled to mean FDG uptake in brain parenchyma as reference region, because this had resulted in better prognostic power for prediction of

AD dementia in mild cognitive impairment subjects than scaling to the pons in a previous study [23].

The MMSE sum score (standard feature) as well as the other CERAD-plus subscores (additional features) were corrected for age, gender, and education prior to automatic classification. A detailed description of the features is given in Supplementary Material 1. There were 141 features in total, 63 standard features and 78 additional features.

Additionally, a set of expert features was determined in accordance with [33]. It comprised 47 features that are marked by asterisk in Supplementary Material 1.

Standard protocol approvals, registrations, and patient consents

The study was approved by the ethics committee of the state Berlin, Germany (13/0234-EK12). It was registered in the German trials registry which is an approved primary register in the WHO network (DRKS00005041). Written informed consent was obtained from all participants in the study.

Legal capacity to consent in participation in the study was tested by an independent physician based on the standardized criteria described in [34].

RESULTS

Demographic data of the 81 patients are given in Table 1. The patients had been acutely hospitalized because of bone fracture or other injury (37.0%), cardiovascular disease (14.8%), reduced general health (8.6%), stroke (4.9%), pulmonary disease (3.7%), or another cause (30.9%, infection, recurrent diarrhea, gait disturbance, kidney failure, etc.). All patients were recruited in an acute care geriatrics unit. One part of the patients had been directly hospitalized in the geriatrics unit; the other had been first hospitalized in another unit (e.g., orthopedics) from which they were later directly referred to the geriatrics unit. Twenty-four patients (29.6%) had undergone surgery during the current hospital stay prior to enrollment. The number of comorbidities ranged between 5 and 23 (13 on average). The most frequent comorbidities were hypertension (in 84.0% of the patients), lipometabolic disorder (61.7%), depression (53.1%), thyroid disease (38.3%), diabetes mellitus (33.3%), chronic obstructive pulmonary disease/asthma (19.8%), stroke (18.5%), and myocardial infarction (17.3%). One third of the patients had manifested a delirium during their current hospital

Table 1
Patient characteristics

Characteristic	non-neurodegenerative (n = 15)	AD (n = 17)	CVD (n = 23)	MD (n = 26)
Age [years]	82.13 ± 5.48	79.18 ± 3.84	80.52 ± 6.20	83.88 ± 4.99
Gender [number of females]	11 (73.3%)	13 (76.5%)	13 (56.5%)	18 (69.2%)
Education [years]	10.97 ± 2.74	12.41 ± 3.54	12.61 ± 3.71	11.96 ± 3.57
MMSE [z-score]	-2.73 ± 1.74	-3.99 ± 1.37	-3.18 ± 1.70	-4.01 ± 1.36

Numbers are given as mean ± standard deviation (AD, Alzheimer's disease; CVD, cerebrovascular disease; MD, mixed etiology; MMSE, Mini-Mental State Examination).

stay, which, however, at time of enrollment was in remission so that delirium was not considered a likely cause of the cognitive impairment. This was confirmed, after enrollment, by the Nursing Delirium Score [35] and the Delirium Rating Scale [36]. Medication included on average 9 different drugs (range: 1–16). Severity of cognitive impairment according to the CDR score was mild cognitive impairment (CDR = 0.5) in 65.4% of the patients and mild-to-moderate dementia (CDR > 0.5) in 32.1% (Supplementary Material 1). Two patients (2.5%) presented with CDR = 0.

Etiologic diagnosis by the team of experts was non-neurodegenerative etiology in 15 patients (18.5%), AD in 17 patients (21.0%), CVD in 23 patients (28.4%), and MD in 26 patients (32.1%).

Full description of the patients, separated according to etiologic diagnosis and including all considered features, is given in Supplementary Material 1.

Single features

Automatic classification results based on single features are shown in Fig. 2 and are listed in Supplementary Material 1 for each single feature. Best agreement of automatic single feature classification with the expert etiologic diagnosis was obtained with the scaled FDG uptake within a predefined AD mask (45.4%, chance level: 25%). Among the NP features (standard and additional), the BEHAVE score (34.7%) and the CERAD subtests naming animals (36.5%), wordlist recognition and recall (34.4%, 33.7%), and figures recall (32.1%) were most discriminative. Among the vascularMRI features, highest accuracy was achieved by severity of white matter hyperintensities (44.6%), existence (yes/no) of periventricular white matter hyperintensities (40.4%), total number of lacunes (40.7%), and existence (yes/no) of chronic lacunes in bilateral thalamus (39.4%). Among the volumetric MRI features, grey matter volume in left and right lateral temporal lobe provided the highest discrimination power (36.1%, 34.0%; Fig. 3). Prior surgery and delirium,

conditions that might have been expected to affect the prediction performance, resulted in single feature classification accuracy that was not statistically different from chance level.

Single modality versus all modalities

Classification results for single modalities and all modalities together are shown in Fig. 4. Agreement of the automatic 4-class classification (chance level 25%) with the etiologic expert diagnosis ranged between 25.5% (clinical, $p = 0.42$) and 45.1% (vascularMRI, $p \leq 0.001$). Next to best agreement was obtained with PET (44.7%, $p \leq 0.001$) and NP (standardNP+additionalNP, 42.6%, $p \leq 0.001$). Integration of all modalities provided a considerable increase in the accuracy of automatic 4-class etiologic diagnosis to 76.4% ($p \leq 0.001$).

For detection of AD (among all subjects), NP, volumetricMRI and PET features were particularly important (accuracy for detection of AD: 45.1%, 30.8%, and 32.4%).

Standard modalities versus standard+additional modalities

When automatic classification was allowed to use only the standard features, accuracy of the 4-class etiologic diagnosis declined to 31.4% ($p = 0.100$), from 76.4% with all features included, i.e., both standard and additional features.

In the detection of a single etiology, restriction to standard features resulted in a decrease of accuracy from 87.8 to 22.4%, from 80.0 to 38.8%, from 52.8 to 28.7%, and from 85.1 to 35.6% for non-neurodegenerative etiology, AD, CVD, and MD, respectively.

Expert features

When the classification analysis was limited to the expert features ($n = 47$), an accuracy of 45.6% ($p \leq 0.001$) was achieved.

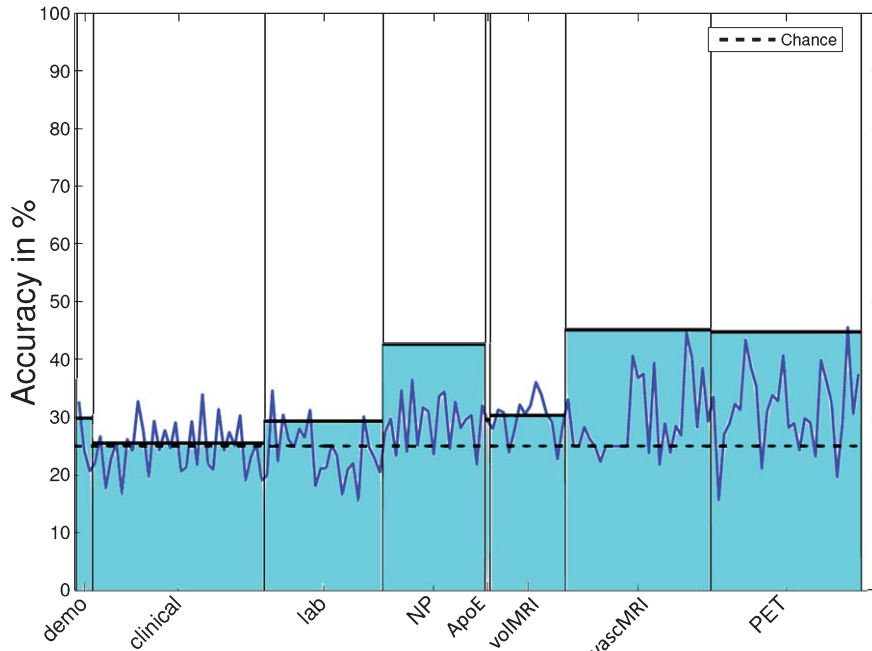


Fig. 2. Accuracy of automatic 4-class etiological diagnosis for each single feature (black line) and for all features of a given modality together (grey batches). (demo, demographical features; clinical, clinical features; lab, standard blood and urine laboratory features; NP, standard and additional neuropsychological test scores; ApoE, ApoE genotype; volMRI, grey matter volume in different predefined brain regions; vascMRI, visual MRI-based scores for vascular pathology; PET, scaled FDG uptake in different predefined brain regions).

Detection of neurodegeneration and cerebrovascular disease

Most successful in detecting neurodegenerative disease (AD or MD versus all others) was PET (85.8% accuracy, $p \leq 0.001$). Significant classification accuracy was also achieved by the modalities NP, clinical data and genotype (65.4%, 62.2%, and 61.5%, $p \leq 0.05$). Integrating all modalities did not improve detection of neurodegeneration compared to PET alone (85.1%, $p \leq 0.001$).

Most successful in detecting cerebrovascular disease (CVD or MD versus all others) was vascularMRI with an accuracy of 91.2% ($p \leq 0.001$). Significant classification accuracy was also achieved by PET, NP, and volumetricMRI (66.5%, 64.5%, and 61.1%, $p \leq 0.05$). Integrating all modalities did not improve detection of cerebrovascular disease compared to vascularMRI alone (91.2%, $p \leq 0.001$).

Chance level in these two analyses was 50 % (not 25 % as in the 4-class case).

DISCUSSION

Primary goal of this study was to evaluate the utility of biomarkers derived from structural and metabolic

brain imaging (MRI and FDG-PET) in addition to detailed neuropsychological testing for the etiologic diagnosis of clinically uncertain cognitive impairment (CUCI) in elderly patients hospitalized in a geriatrics unit for an acute or subacute cause without history of cognitive impairment prior to hospitalization. There is a considerable lack of studies on the performance of biomarkers in this particular clinical setting even though cognitive decline and dementia are more common in elderly inpatients than in subjects of the same age in the community [37, 38]. The lack of data on the use of biomarkers in this setting is also reflected in current diagnostic AD criteria. The NIA-AA criteria require gradual onset of cognitive decline over months to years for the diagnosis of probable AD [7], cognitive decline occurring within days to weeks during hospitalization excludes the diagnosis of probable AD even if there is established biomarker-based evidence of the AD pathophysiological process.

We hypothesize that the diagnostic benefit of using biomarkers is even larger in this specific geriatric inpatient setting than in the typical outpatient memory clinic setting. It is well known that patients in acute hospitals may perform poorly in cognitive testing for other reasons than neurodegenerative or

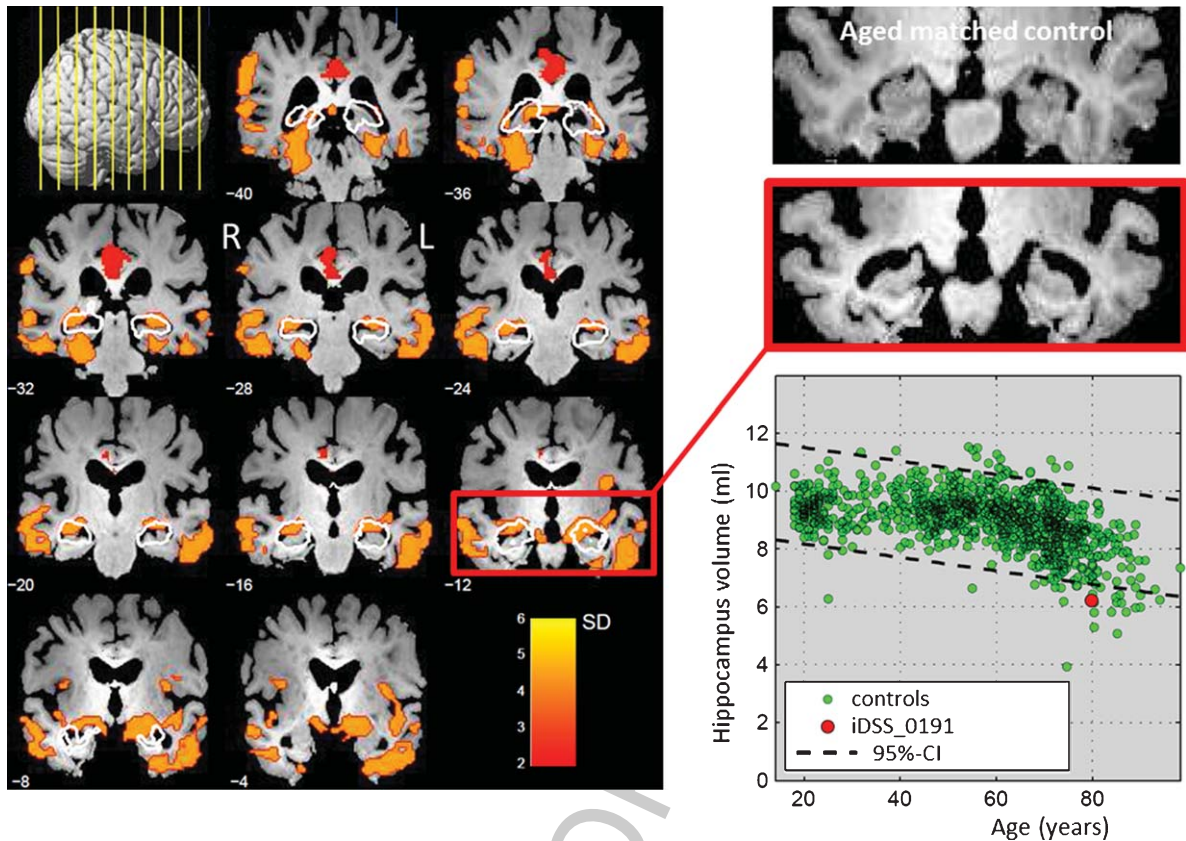


Fig. 3. Quantitative characterization of temporal atrophy on MPRAGE MRI. Left: statistical parametric map from voxel-based morphometry superimposed to coronal slices of the patient's stereotactically normalized MPRAGE. Top right: Scheltens window [47] of the patient (framed in red) and an age-matched healthy control subject (without frame). Bottom right: Bilateral hippocampus volume (adjusted for total intracranial volume) estimated as described previously [31] of the patient (red dot) and healthy control subjects (green dots). Hippocampus volume is plotted versus age. The images are from a 79 years old female categorized as mixed etiology (iDSS_0191).

cerebrovascular disease such as acute illness, pain, lethargy, sleep deprivation, medication, depression, anxiety, or simply not wishing to engage with testing [39]. Therefore, standard diagnostic workup including neuropsychological testing but no biomarkers might not be reliable enough for etiologic diagnosis of cognitive impairment in the acute geriatric hospital setting [40].

This study included 'unclear' geriatric inpatients with complex clinical presentations. The patients were relatively old, 81.6 ± 5.5 years on average (9 patients ≥ 90 years), had been hospitalized for various acute or subacute severe syndromes, and presented with many comorbidities, considerable pain, and extensive medication. Only a small fraction of the participants would have fulfilled eligibility criteria of typical studies on biomarkers for the diagnosis of cognitive impairment. Prior surgery and delirium during the current hospital stay had no significant impact on the etiological diagnosis in this patient sample.

The complementary benefit from the additional diagnostic procedures was tested by automatic etiological single subject classification by a machine learning method either using all available data, i.e., from both standard and additional procedures together, or using the data from standard clinical workup only. The rationale was that the difference in the accuracy of automatic classification between these two scenarios is a measure of the complementary benefit from the additional diagnostic procedures.

Single features and single modalities

Among the 141 single features, scaled FDG uptake in a predefined AD mask [23] provided the highest accuracy for the automatic 4-class etiological diagnosis (45.4%). Among the NP features, especially naming, recognition and recall tests were relevant for class discrimination, in agreement with previous

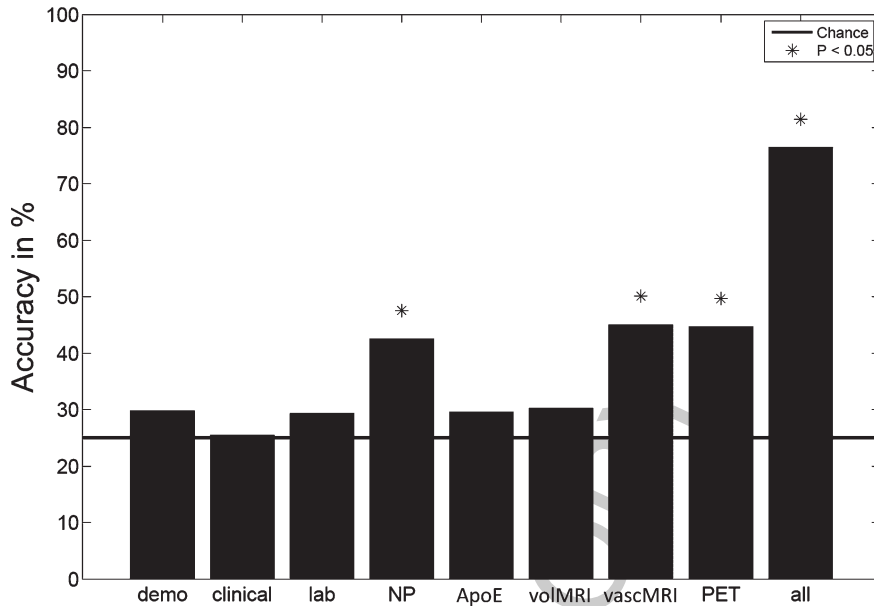


Fig. 4. Accuracy of automatic 4-class etiological diagnosis for each single modality and combination of all modalities. The chance level is at 25%. Modalities with a permutation p -value below 0.05 for being better than chance are marked by a *. (demo, demographical features; clinical, clinical features; lab, standard blood and urine laboratory features; NP, standard and additional neuropsychological test scores; ApoE, ApoE genotype; volMRI, grey matter volume in different predefined brain regions; vascMRI, visual MRI-based scores for vascular pathology; PET, scaled FDG uptake in different predefined brain regions; all, all features together).

studies [33]. Among the 8 single modalities, the combination of various vascular MRI scores provided the best discrimination power (45.1%). However, accuracies below 50% achieved by single features or single modalities are far too low to be clinically useful. Thus, integration of features and / or modalities is mandatory.

Standard versus additional modalities

The standard diagnostic modalities (demographical data, clinical data, standard blood and urine laboratory tests, and standard neuropsychological testing) are routinely used in the diagnostic workup of elderly patients hospitalized in a geriatrics ward. Their primary purpose is to detect and characterize the severity of cognitive impairment and to detect or exclude treatable causes. Standard diagnostic modalities may also allow an etiologic diagnosis, particularly in patients with typical presentation. However, the fraction of cognitive impairment that is clinically uncertain (CUCI) is particularly large in the geriatric inpatient setting. The present results suggest that the standard diagnostic modalities alone provide only little power for etiological diagnosis in this setting (accuracy of automatic classification based on standard modalities alone was 31.4%, not

significantly different from the chance level of 25%, $p=0.100$). In contrast, the present results suggest a particularly large add-on value of the additional diagnostic procedures in this setting, in line with the primary hypothesis of the study.

Expert features

The acquisition of 141 features seems not feasible in everyday clinical routine. We therefore restricted the analysis to the 47 (apparently) most important features based on expert opinion. Notably, this resulted in a decrease of accuracy from 76.4% to 45.6% arguing for the strength of boosting algorithms to extract information from large sets of noisy features.

Detection of neurodegeneration and cerebrovascular disease

Whereas FDG-PET features were most successful in detecting neurodegeneration, vascular features derived from MRI were best in identifying cerebrovascular disease, as was to be expected. In both cases, integration of other modalities did not improve detection accuracy, suggesting that other modalities did not provide complementary information. For both pathologies, also NP resulted in above

chance classification accuracy. Notably, volumetric MRI reached the level of statistical significance only for identification of cerebrovascular disease but not neurodegeneration.

Structural MRI versus FDG-PET

Concerning the relative contribution of structural MRI and brain FDG-PET, the present results suggest that detailed MRI-based characterization of cerebrovascular disease and FDG-PET assessment for regional detection/exclusion of AD-typical synaptic dysfunction both contribute independently and substantially to the etiological diagnosis (previous paragraph). This finding is in agreement with previous studies on multimodal diagnoses of AD and CVD in clinical settings [41]. It is also in line with what is known to be the strengths and weaknesses of both imaging modalities.

Limitations

The following limitations should be noted. First, and most important, the sample size of this study is relatively small. However, in order to minimize small sample bias [42], the state-of-the-art machine learning algorithm RUSBoost was used that has been designed to deal with the challenges of small, incomplete and unbalanced data sets.

Second, the clinical usefulness of an etiological diagnosis with 76% accuracy as achieved by automatic classification integrating all features is questionable, although it should be noted that this accuracy level is for automatic classification according to four etiologic diagnoses for which the chance level is 25% (not 50% as in the case of two diagnostic categories). Nevertheless, automatic classification should aim at higher discriminatory power. Integrating information about amyloid pathology in the brain, either from CSF or from amyloid-PET, is clearly a promising approach, although the incremental benefit from amyloid markers is still not very clear for “difficult” settings. In the present study, CSF analysis for amyloid- β (and tau) proteins was performed. However, only 10 of the 81 patients consented in lumbar puncture so that CSF was not included in the analyses.

Third, a general assumption of automatic single-label classification is that all classes are mutually exclusive. Therefore, mixed disease (MD) was considered a separate etiologic disease class, independent from AD and CVD, although MD might also be considered a mix of AD and CVD. The rationale

for this was that the inclusion of mixed classes as extra classes often provides better performance in automatic classification than multi-label approaches [43].

Fourth, collinearity of features was not taken into account. Although most classification algorithms have problems with large and noisy data sets comprising also uninformative and redundant data, a recent review concluded that most machine learning methods (including decision trees and boosting algorithms) outperform the general linear model in the case of collinearity [44]. The same review also concluded that in terms of accuracy it does not make a big difference whether one ‘ignores’ collinearity in the data or applies collinearity reduction methods such as the variance inflation factor. Boosting algorithms alleviate the problem of multicollinearity by shrinkage of effect estimates similar to penalized regression approaches [45]. The addition of an apparently redundant variable can actually lead to better classification performances [46].

Finally, acquisition and processing of data in this study was rather extensive and complex. For multimodal diagnosis, including automatic classification, to be feasible in clinical routine, the number of features has to be strongly reduced. However, the limited set of 47 expert features resulted in a much lower accuracy than the whole set. To find the most powerful diagnosis algorithm, future studies should evaluate different combinations of features, automatic feature selection methods (such as principal component analysis) and classification methods. In addition, ways of automatizing the extraction of features from the image data and fully automated classification pipelines have to be developed for multimodal classification.

ACKNOWLEDGMENTS

This work was supported by the Regional Development Fund of the European Union (10153407, 10153971, 10153458, 10153460, 10153461, 10153462, 10153463), the Bernstein Computational Program of the German Federal Ministry of Education and Research (01GQ1001C, 01GQ0851, GRK 1589/1), and Philips Research.

We thank Stephanie Liese, Hans-Joachim Igel and Julie O’Sullivan (all Evangelisches Geriatriezentrum Berlin, Berlin, Germany) for their sustained commitment to recruitment and supporting the patients during their participation in the study. We also thank (in alphabetical order, all Berlin, Germany)

Dr. Sibylle Gebauer (Dominikus Krankenhaus), Dr. Eric Hilf (Sana Klinikum Lichtenberg), Dr. Rainer Koch (Krankenhaus Hedwigshöhe), Dr. Martin K. Kuhlmann (Vivantes Klinikum im Friedrichshain), Dr. Gernot Lämmler and Dr. Elsiabeth Stechl (both Evangelisches Geriatriezentrum Berlin), Dr. Andreas Schwitzke (Vitanas Klinik für Geriatrie), Dr. Christian Stofft (Sankt Gertrauden Krankenhaus), Prof. Dr. Hans-Peter Thomas (Evangelisches Krankenhaus Königin Elisabeth Herzberge), Dr. Martin Warnach (Wichern-Krankenhaus), and their staff for supporting patient recruitment.

Authors' disclosures available online (<http://j-alz.com/manuscript-disclosures/16-0380r1>).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/jad-160380>.

REFERENCES

- [1] Fried TR, Bradley EH, Towle VR, Allore H (2002) Understanding the treatment preferences of seriously ill patients. *N Engl J Med* **346**, 1061-1066.
- [2] Alzheimer's Association (2014) Alzheimer's Association Report: 2014 Alzheimer's Disease Facts and figures. *Alzheimers Dement* **10**, e47-e92.
- [3] (2009) Deutsche Gesellschaft für Neurologie (DGN) und Deutsche Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde e. V. (DGPPN), AWMF Registernummer 038-013.
- [4] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-944.
- [5] Knopman DS, DeKosky ST, Cummings JL, Chui H, Corey-Bloom J, Relkin N, Small GW, Miller B, Stevens JC (2001) Practice parameter: Diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology* **56**, 1143-1153.
- [6] Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, DeKosky ST, Gauthier S, Selkoe D, Bateman R, Cappa S, Crutch S, Engelborghs S, Frisoni GB, Fox NC, Galasko D, Habert MO, Jicha GA, Nordberg A, Pasquier F, Rabinovici G, Robert P, Rowe C, Salloway S, Sarazin M, Epelbaum S, de Souza LC, Vellas B, Visser PJ, Schneider L, Stern Y, Scheltens P, Cummings JL (2014) Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurol* **13**, 614-629.
- [7] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) 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* **7**, 263-269.
- [8] Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, Jack CR Jr, Jagust WJ, Shaw LM, Toga AW, Trojanowski JQ, Weiner MW (2010) Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology* **74**, 201-209.
- [9] Weise D, Tiepolt S, Awissus C, Hoffmann KT, Lobsien D, Kaiser T, Barthel H, Sabri O, Gertz HJ (2015) Critical comparison of different biomarkers for Alzheimer's disease in a clinical setting. *J Alzheimers Dis* **48**, 425-432.
- [10] Brindle N, Holmes J (2005) Capacity and coercion: Dilemmas in the discharge of older people with dementia from general hospital settings. *Age Ageing* **34**, 16-20.
- [11] Frisoni GB, Fox NC, Jack CR Jr, Scheltens P, Thompson PM (2010) The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol* **6**, 67-77.
- [12] Silverman DH, Small GW, Chang CY, Lu CS, Kung De Aburto MA, Chen W, Czernin J, Rapoport SI, Pietrini P, Alexander GE, Schapiro MB, Jagust WJ, Hoffman JM, Welsh-Bohmer KA, Alavi A, Clark CM, Salmon E, de Leon MJ, Mielke R, Cummings JL, Kowell AP, Gambhir SS, Hoh CK, Phelps ME (2001) Positron emission tomography in evaluation of dementia: Regional brain metabolism and long-term outcome. *JAMA* **286**, 2120-2127.
- [13] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.
- [14] Kalbe E, Kessler J, Calabrese P, Smith R, Passmore AP, Brand M, Bullock R (2004) DemTect: A new, sensitive cognitive screening test to support the diagnosis of mild cognitive impairment and early dementia. *Int J Geriatr Psychiatry* **19**, 136-143.
- [15] Morris JC (1993) The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* **43**, 2412-2414.
- [16] Reisberg B, Borenstein J, Salob SP, Ferris SH, Franssen E, Georgotas A (1987) Behavioral symptoms in Alzheimer's disease: Phenomenology and treatment. *J Clin Psychiatry* **48 Suppl.** 9-15.
- [17] Leshner EL, Berryhill JS (1994) Validation of the Geriatric Depression Scale-Short Form among inpatients. *J Clin Psychol* **50**, 256-260.
- [18] Monsch AU, Kressig RW (2010) Specific care program for the older adults: Memory Clinics. *Eur Geriatr Med* **1**, 128-131.
- [19] Lehl S, Merz J, Burkhard G, Fischer S (1991) Mehrfachwahl-Wortschatz-Intelligenztest, Perimed, Erlangen.
- [20] Montgomery SA, Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* **134**, 382-389.
- [21] Jorm AF (1994) A short-form of the Informant Questionnaire on Cognitive Decline in the Elderly (Iqcode) - development and cross-validation. *Psychol Med* **24**, 145-153.
- [22] Varrone A, Asenbaum S, Vander Borgh T, Booi J, Nobili F, Nagren K, Darcourt J, Kapucu OL, Tatsch K, Bartenstein P, Van Laere K, European Association of Nuclear Medicine Neuroimaging, Committee (2009) EANM procedure guidelines for PET brain imaging using [¹⁸F]FDG, version 2. *Eur J Nucl Med Mol Imaging* **36**, 2103-2110.

- [23] Lange C, Suppa P, Frings L, Brenner W, Spies L, Buchert R (2015) Optimization of statistical single subject analysis of brain FDG PET for the prognosis of mild cognitive impairment-to-Alzheimer's disease conversion. *J Alzheimers Dis* **49**, 945-959.
- [24] Andersson JL, Vagnhammar BE, Schneider H (1995) Accurate attenuation correction despite movement during PET imaging. *J Nucl Med* **36**, 670-678.
- [25] Roman GC, Tatemichi TK, Erkinjuntti T, Cummings JL, Masdeu JC, Garcia JA, Amaducci L, Orgogozo JM, Brun A, Hofman A (1993) Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology* **43**, 250-260.
- [26] Brenowitz WD, Monsell SE, Schmitt FA, Kukull WA, Nelson PT (2014) Hippocampal sclerosis of aging is a key Alzheimer's disease mimic: Clinical-pathologic correlations and comparisons with both Alzheimer's disease and non-tauopathic frontotemporal lobar degeneration. *J Alzheimers Dis* **39**, 691-702.
- [27] Nelson PT, Smith CD, Abner EL, Wilfred BJ, Wang WX, Neltner JH, Baker M, Fardo DW, Kryscio RJ, Scheff SW, Jicha GA, Jellinger KA, Van Eldik LJ, Schmitt FA (2013) Hippocampal sclerosis of aging, a prevalent and high-morbidity brain disease. *Acta Neuropathol* **126**, 161-177.
- [28] Haller S, Lovblad K-O, Giannakopoulos P, Van De Ville D (2014) Multivariate pattern recognition for diagnosis and prognosis in clinical neuroimaging: State of the art, current challenges and future trends. *Brain Topogr* **27**, 329-337.
- [29] Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A (2010) RUSBoost: A hybrid approach to alleviating class imbalance. *IEEE Trans Syst Man Cybern A Syst Hum* **40**, 185-197.
- [30] Bishop CM (2007) Pattern Recognition and Machine Learning. Springer.
- [31] Suppa P, Anker U, Spies L, Bopp I, Ruegger-Frey B, Klaghofer R, Gocke C, Hampel H, Beck S, Buchert R (2015) Fully automated atlas-based hippocampal volumetry for detection of Alzheimer's disease in a memory clinic setting. *J Alzheimers Dis* **44**, 183-193.
- [32] Suppa P, Hampel H, Spies L, Fiebach JB, Dubois B, Buchert R (2015) Fully automated atlas-based hippocampus volumetry for clinical routine: Validation in subjects with mild cognitive impairment from the ADNI Cohort. *J Alzheimers Dis* **46**, 199-209.
- [33] Ritter K, Schumacher J, Weygandt M, Buchert R, Allefeld C, Haynes JD (2015) Multimodal prediction of conversion to Alzheimer's disease based on incomplete biomarkers. *Alzheimers Dement (Amst)* **1**, 206-215.
- [34] Bühler E, Kren R, Stolz K (2012) *Betreuungsrecht und Patientenverfügungen: Praktische Informationen für Ärzte und Interessierte*. Springer Medizin Verlag GmbH.
- [35] Gaudreau JD, Gagnon P, Harel F, Tremblay A, Roy MA (2005) Fast, systematic, and continuous delirium assessment in hospitalized patients: The nursing delirium screening scale. *J Pain Symptom Manage* **29**, 368-375.
- [36] Trzepacz PT, Mulsant BH, Dew MA, Pasternak R, Sweet RA, Zubenko GS (1998) Is delirium different when it occurs in dementia? A study using the delirium rating scale. *J Neuropsychiatry Clin Neurosci* **10**, 199-204.
- [37] Matthews FE, Arthur A, Barnes LE, Bond J, Jagger C, Robinson L, Brayne C, Medical Research Council Cognitive F, Ageing C (2013) A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: Results of the Cognitive Function and Ageing Study I and II. *Lancet* **382**, 1405-1412.
- [38] Sampson EL, Blanchard MR, Jones L, Tookman A, King M (2009) Dementia in the acute hospital: Prospective cohort study of prevalence and mortality. *Br J Psychiatry* **195**, 61-66.
- [39] Mathews SB, Arnold SE, Epperson CN (2014) Hospitalization and cognitive decline: Can the nature of the relationship be deciphered? *Am J Geriatr Psychiatry* **22**, 465-480.
- [40] Shenkin SD, Russ TC, Ryan TM, MacLulich AM (2014) Screening for dementia and other causes of cognitive impairment in general hospital in-patients. *Age Ageing* **43**, 166-168.
- [41] Wallin A, Nordlund A, Jonsson M, Blennow K, Zetterberg H, Ohrfelt A, Stalhammar J, Eckerstrom M, Carlsson M, Olsson E, Gotthlin M, Svensson J, Rolstad S, Eckerstrom C, Bjerke M (2016) Alzheimer's disease-subcortical vascular disease spectrum in a hospital-based setting: Overview of results from the Gothenburg MCI and dementia studies. *J Cereb Blood Flow Metab* **36**, 95-113.
- [42] Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR (2013) Power failure: Why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* **14**, 365-376.
- [43] Tsoumakas G, Katakis I (2007) Multi-label classification: An overview. *Int J Data Warehousing Min* **2007**, 1-13.
- [44] Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Marquez JRG, Gruber B, Lafourcade B, Leitao PJ, Munkemüller T, McClean C, Osborne PE, Reineking B, Schroder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27-46.
- [45] Mayr A, Binder H, Gefeller O, Schmid M (2014) The evolution of boosting algorithms from machine learning to statistical modelling. *Methods Inf Med* **53**, 419-427.
- [46] Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* **3**, 1157-1182.
- [47] Scheltens P, Leys D, Barkhof F, Huglo D, Weinstein HC, Vermersch P, Kuiper M, Steinling M, Wolters EC, Valk J (1992) Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: Diagnostic value and neuropsychological correlates. *J Neurol Neurosurg Psychiatry* **55**, 967-972.

Appendix: Eligibility Criteria

Inclusion criteria

- the patient is hospitalized in a geriatrics ward with expected remaining duration of the stay \geq 1 week
- clinically uncertain cognitive impairment (CUCI) after standard diagnostic workup including patient history, physical/neurological examination, standard neuropsychological testing (including mini-mental state examination, MMSE, and optionally geriatric depression scale, GDS, and clock drawing test), and standard blood/urine laboratory tests. No further dementia-specific tests performed prior to inclusion
- suspected etiology AD or CVD or MD
- 'unclear case', i.e., the cause of the CUCI is sufficiently unclear so that additional diagnostic evaluation including biomarkers appears useful for each individual participant
- general health condition allows participation in the trial (this excludes acute infections, instable somatic conditions, instable trauma, nosocomial resistant pathogens . . .)
- age \geq 65 years
- a relative of the patient is available to support the patient's participation and for external assessment
- MMSE \geq 18 (rather low threshold to account for sensoric deprivation)
- written informed consent given

Exclusion criteria

- history of dementia prior to hospitalization
- history of anti-dementia therapy

- expected survival $<$ 3 years
- suspicion of neurodegenerative disease other than AD (frontotemporal lobar degeneration, Morbus Parkinson (with dementia)/Lewy body disease, atypical Parkinsonian syndromes)
- strong indication of secondary dementia (delirium, hypothyreosis, vitamin B12 deficiency, infection, normal pressure hydrocephalus, brain tumor) which would violate the 'unclear case' inclusion criterion
- acute psychiatric disease which interferes with the participation (e.g., schizophrenia, manic depression)
- history of brain trauma, if the cognitive impairment suddenly started at the time of the trauma
- large stroke, but only if it is the cause of the cognitive impairment with high probability
- medication with strong effect on cognition and/or cerebral glucose metabolism (e.g., anti-convulsants)
- MRI-specific exclusion criteria
- low compliance expected
- lack of legal capacity to consent in the participation
- arrested by court order
- lack of consent in archiving and transfer of pseudonymized data
- pregnancy and/or breast feeding
- participation in another medicinal product trial during the last 3 months
- participation in another clinical study with radiation exposure in the last 3 months