



## OPEN ACCESS

## EDITED BY

Sotirios Bisdas,  
University College London, United Kingdom

## REVIEWED BY

Swati Goyal,  
Gandhi Medical College Bhopal, India  
Benedetta Tafuri,  
University of Salento, Italy  
Changhong Jing,  
Hong Kong Polytechnic University, Hong  
Kong SAR, China

## \*CORRESPONDENCE

S. F. Calloni  
✉ calloni.sonia@hsr.it

<sup>†</sup>These authors have contributed equally to  
this work and share first authorship

RECEIVED 16 July 2025

REVISED 09 December 2025

ACCEPTED 05 January 2026

PUBLISHED 27 January 2026

## CITATION

Calloni SF, Diena A, Agazzi GM, Zavarella M,  
Vezzulli PQ, Cecchetti G, Spinelli EG,  
Rugarli G, Ghirelli A, Magnani G, Caso F,  
van Loon A, Agosta F, Filippi M and Falini A  
(2026) MRI approach to the patient with  
suspected dementia: artificial intelligence  
techniques and semi-quantitative rating scales  
compared.  
Front. Radiol. 6:1667306.  
doi: 10.3389/fradi.2026.1667306

## COPYRIGHT

© 2026 Calloni, Diena, Agazzi, Zavarella,  
Vezzulli, Cecchetti, Spinelli, Rugarli, Ghirelli,  
Magnani, Caso, van Loon, Agosta, Filippi and  
Falini. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# MRI approach to the patient with suspected dementia: artificial intelligence techniques and semi-quantitative rating scales compared

S. F. Calloni<sup>1\*†</sup>, A. Diena<sup>1†</sup>, G. M. Agazzi<sup>1</sup>, M. Zavarella<sup>2</sup>,  
P. Q. Vezzulli<sup>1</sup>, G. Cecchetti<sup>2,3,4</sup>, E. G. Spinelli<sup>3,4,5</sup>, G. Rugarli<sup>2,4,5</sup>,  
A. Ghirelli<sup>2,4,5</sup>, G. Magnani<sup>6</sup>, F. Caso<sup>6</sup>, A. van Loon<sup>7</sup>, F. Agosta<sup>4,5,6,8</sup>,  
M. Filippi<sup>2,3,4,5,6,8,9</sup> and A. Falini<sup>1,4</sup>

<sup>1</sup>Neuroradiology Unit and Cermac, IRCCS Ospedale San Raffaele, Milan, Italy, <sup>2</sup>Center for Alzheimer's Disease and Related Disorders (CARD), Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy, <sup>3</sup>Neurophysiology Service, IRCCS San Raffaele Scientific Institute, Milan, Italy, <sup>4</sup>Vita-Salute San Raffaele University, Milan, Italy, <sup>5</sup>Neuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy, <sup>6</sup>Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy, <sup>7</sup>DeepHealth, Boston, MA, United States, <sup>8</sup>Neurotech Hub, Vita-Salute San Raffaele University, Milan, Italy, <sup>9</sup>Neurorehabilitation Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy

**Objectives:** To assess the reliability of semi-quantitative and AI-based quantitative brain volume evaluation (Quantib<sup>®</sup> ND) in predicting clinical diagnosis in patients with suspected neurodegenerative diseases undergoing initial 1.5 T MRI. Additionally, to analyze the frequency of lobar microbleeds (MBs) at diagnosis.

**Methods:** Two neuroradiologists (2 vs. 10 years' experience), blinded to diagnosis, independently evaluated brain atrophy on 3D-T1 images of 133 subjects using Scheltens, Koedam, and Kipps scales. Automated volumetric analysis was performed using Quantib<sup>®</sup> ND. SWI images were assessed by one neuroradiologist to classify MBs as cortical, juxtacortical, subcortical, or deep. Inter-observer agreement was measured using intraclass correlation coefficients (ICC); correlation with Quantib<sup>®</sup> ND was analyzed using Spearman's coefficient. Cohen's Kappa assessed agreement with clinical diagnosis.

**Results:** Good inter-observer agreement was observed for the MTA scale (ICC 0.86 right, 0.82 left) and Kipps scale (ICC 0.76), with moderate concordance for Koedam (ICC 0.66). Frontal and posterior temporal Kipps subregions had good concordance (ICC 0.77, 0.79), while anterior temporal showed poor agreement (ICC 0.59). Diagnostic accuracy was moderate across observers and Quantib<sup>®</sup> ND. Observer 1 showed 77% sensitivity, 51% specificity; observer 2 had 79% sensitivity, 62% specificity; Quantib<sup>®</sup> ND reached 56% sensitivity, 74% specificity. Patients exhibited significantly more lobar MBs than non-dementia patients ( $\chi^2$   $p = 0.04$ ).

**Conclusions:** Semi-quantitative visual scales proved effective and sensitive for detecting brain atrophy, showing good concordance with automated volumetric data. While AI-based quantification demonstrated higher specificity, visual assessment remained more sensitive. Lobar MBs were more frequent in neurodegenerative cases.

## KEYWORDS

artificial intelligence, brain volumetry, dementia, magnetic resonance imaging, microbleeds, Quantib ND, visual rating scales

## 1 Introduction

Dementia affects approximately 55 million people worldwide, with projections indicating that this number could nearly triple by 2050 (1). Alzheimer's disease (AD) and vascular dementia are the leading causes, accounting for around 80% of global dementia cases, with AD responsible for 50%–60% of those (2). As imaging technologies advance, neuroradiology's role in diagnosing cognitive impairment has evolved significantly. What was once primarily focused on excluding diseases is now a crucial tool for clinical diagnosis, subtyping, and ongoing monitoring of neurodegenerative conditions (3). In particular, the role of magnetic resonance imaging (MRI) has expanded as a vital tool for dementia diagnosis. MRI requests are on the rise, especially in centers specializing in neurodegenerative diseases, with the most common field strength being 1.5 Tesla. MRI allows for the assessment of cortical atrophy through visual analysis using semi-quantitative rating scales, which remain highly effective and widely employed in both clinical practice and research for identifying characteristic atrophy patterns in dementia (4). While the quantitative approach is still considered complementary, it promises to enhance the accuracy of diagnosing dementia-related conditions (5, 6). In recent years, artificial intelligence (AI) and machine learning have profoundly advanced neuroimaging-based dementia research. Deep learning architectures—particularly convolutional neural networks (CNNs) and transformer-based models—have achieved high accuracy in detecting and classifying Alzheimer's disease and related dementias from MRI data (7, 8). These algorithms can automatically extract disease-relevant imaging biomarkers from large-scale datasets, often reaching diagnostic performance comparable to expert neuroradiologists.

Nevertheless, these research-oriented AI models remain limited in clinical translation due to dataset heterogeneity, lack of interpretability, and absence of regulatory approval. In contrast, visual rating scales—such as Scheltens' medial temporal atrophy (MTA), Koedam's posterior atrophy, and Kipps' lobar atrophy—are well-established, reproducible tools that remain central to everyday dementia assessment but are inherently observer-dependent. Clinically validated AI-based volumetric software, such as Quantib<sup>®</sup> ND, offers a practical bridge between these two approaches by providing standardized, quantitative measures of brain atrophy within regulated diagnostic workflows. However, few studies (9) have directly compared the reliability and diagnostic performance of such approved AI tools with traditional visual scales under real-world 1.5 T MRI conditions and across readers with differing levels of expertise—a gap that the present study aims to address.

A key concept in the treatment of neurodegenerative diseases, similar to ischemic stroke, is “time is brain.” Early intervention with disease-modifying therapies for Alzheimer's disease could

significantly influence the course of the illness. In this context, integrating the detection of microbleeds (MBs) via susceptibility-weighted imaging (SWI) sequences into the diagnostic process is becoming increasingly important. Microbleeds, particularly when more than four on T2\*-GE images, may be linked to adverse side effects (ARIA-A and ARIA-E) in patients receiving newer immunobiological therapies (10).

This study aims to address the following four objectives:

1. Assess the reliability of semi-quantitative evaluation in predicting clinical diagnoses for patients undergoing initial 1.5 Tesla MRI scans with suspected neurodegenerative disease, and compare the performance of two experts.
2. Explore the correlation between semi-quantitative assessments and quantitative cerebral lobar volumetry processed using commercial AI software (Quantib<sup>®</sup> ND).
3. Evaluate the difference in diagnostic accuracy between semi-quantitative and quantitative methods in detecting cerebral atrophy indicative of dementia.
4. Investigate whether there is a correlation between the number of lobar microbleeds and the clinical diagnosis of dementia.

## 2 Materials and methods

### 2.1 Participants

The local ethics committee for clinical standards on human trials approved this study. Informed consent was obtained from all participants. For this retrospective observational study, patients who underwent MRI with clinical suspicion of cognitive impairment due to a neurodegenerative disease between December 2021 and June 2024 at the Neuroradiology Operating Unit at IRCCS San Raffaele Hospital in Milan, Italy, were included.

Patients were recruited at the Center for Alzheimer's diseases and related disorders, Unit of Neurology and underwent neurological and neuropsychological examinations as part of the diagnostic work-up at the Neurology Unit. Family history was assessed on the basis of direct interview with patients, caregivers and close relatives. A definite clinical diagnosis of neurodegenerative disease was considered as the diagnostic gold standard. Patients receiving a final diagnosis of neurodegenerative dementia at the Memory Clinic were classified according to current international criteria (e.g., NIA-AA criteria for Alzheimer's disease, Rascovsky et al. criteria for behavioral-variant frontotemporal dementia, Gorno-Tempini et al. criteria for primary progressive aphasia variants, and established consensus criteria for other degenerative syndromes, such as posterior cortical atrophy and dementia with Lewy bodies).

The following were considered exclusion criteria for all subjects enrolled in the study: significant comorbidities or substance abuse that could interfere with cognitive functioning; any major/significant systemic, psychiatric, or neurological disease except cognitive impairment; other causes of focal or diffuse brain damage, including lacunes, extensive ischemic

#### Abbreviations

FTD, Frontotemporal Dementia; PCA, Posterior Cortical Atrophy; MTA, Medial Temporal Lobe Atrophy; GCA, Global Cortical Atrophy; ICC, Intraclass Correlation Coefficient; MBs, Microbleeds.

findings/major strokes on baseline MRI presence of significant motion artifacts on routine MRI.

## 2.2 Imaging protocol

The MRI studies were all acquired with a single scanner, Philips d-Stream 1.5 T (Philips Healthcare, Best, The Netherlands), installed at the Neuroradiology Operating Unit. The MRI protocol used in the acquisitions of all subjects included in the study is shown in [Supplementary Table S1](#).

## 2.3 Analysis of MR imaging

### 2.3.1 Semi-quantitative analysis

MRI images of all subjects were evaluated by two neuroradiologists with mild (observer 1) and high (observer 2) levels of expertise based on their years of neuroradiological experience, (2 and 10 years, respectively).

All 3D T1-weighted images were anonymized for cloud submission to AI brain volumetry quantification software (Quantib<sup>®</sup> ND).

The two observers performed a semi-quantitative assessment of the degree of atrophy on the 3D-T1 images, in a randomized and independent manner, using Scheltens' (values 0-4) Koedam's (values 0-3) and Kipps' (values 0-4) (8-10) visual rating scales, blinded to the clinical diagnosis (11-13).

- Scheltens scale assessment (MTA) was conducted by considering decade-specific cut-offs by calculating the mean of the two sides, specifically the figure is considered pathological in: subject < 65 years if  $MTA \geq 1$ , subject 65-74 years if  $MTA \geq 1.5$ , subject 75-84 years  $MTA \geq 2$ , and subject > 85 years if  $MTA \geq 2$ .

- The Koedam scale assessment was conducted by considering the value of the side with the highest grade, and if this was found to be  $\geq 2$  then the subject was considered pathological.
- Kipps scale evaluation was conducted by considering among each side and region the highest value, if this was  $\geq 2$  the patient is considered pathological.

Subjects with a pathological score in at least one of the three scales were classified as pathological MRI patients, while subjects with none of the three positive scales were considered as negative MRI non-dementia patients.

### 2.3.2 Quantitative analysis

All 3D-T1 images were de-identified and processed using Quantib<sup>®</sup> ND software (<https://www.quantib.com/solutions/quantib-nd>). [Figures 1, 2](#) show an example of a segmentation of a brain T1 3D weighted image of a dementia patient and the relative report, respectively. This software was chosen because of its relative prevalence in centers dedicated to the study of neurodegenerative diseases, because of its easy clinical applicability, because it is methodologically consistent, and not least because it is FDA and CE approved. The software segments and calculates the volumetry of individual brain lobes divided by side (frontal, parietal, temporal, occipital, and hippocampus) ([Figure 1](#)), as well as the total brain volume, identifying volumetry data as pathological if less than the 5 percentile ([Figure 2](#)).

## 2.4 Analysis of microbleeds

All magnetic susceptibility imaging (SWI) was anonymized and evaluated by a single neuroradiologist (observer 1). MBs were counted and classified following a topographical criterion and in accordance with literature (14) into superficial (located solely in the gray matter of the cortex) and juxtacortical (at the interface between cortical gray matter and subcortical white

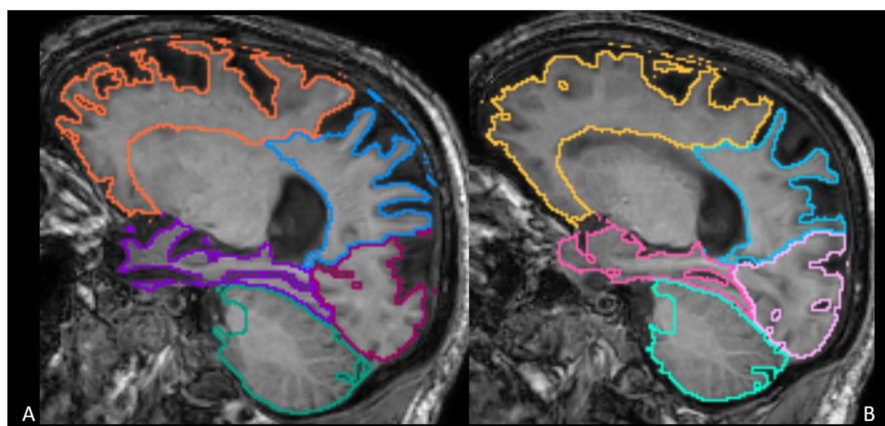


FIGURE 1

The volumetric segmentation map generated by Quantib<sup>®</sup> ND with color-coded anatomical regions and corresponding volumetric data both for right (A) and left side (B).

Brain volumetry				
		Percentile	Volume % of ICV	Volume cm <sup>3</sup>
	CSF	-	26.4	383
	Brain	4.1	73.6	1067
	ICV	-	-	1449

Hippocampus volumetry				
		Percentile	Volume % of ICV	Volume cm <sup>3</sup>
	Total	83.3	0.4	6.0
	Right	98.2	0.2	3.4
	Left	42.2	0.2	2.6

Lobes and cerebellum volumetry				
		Percentile	Volume % of ICV	Volume cm <sup>3</sup>
<b>Frontal</b>				
	Total	7.0	21.6	312
	Right	12.6	10.9	159
	Left	4.4	10.6	154
<b>Occipital</b>				
	Total	0.2	5.7	83
	Right	0.3	2.8	40
	Left	0.6	2.9	43
<b>Parietal</b>				
	Total	2.4	14.0	203
	Right	8.0	7.2	104
	Left	0.9	6.8	99
<b>Temporal</b>				
	Total	2.0	11.7	170
	Right	21.6	6.5	94
	Left	0.2	5.2	76
<b>Cerebellum</b>				
	Total	-	10.0	145
	Right	-	5.0	73
	Left	-	5.0	72

**FIGURE 2**  
 Example of a Quantib® ND atrophy analysis report, illustrating the quantitative percentile rankings of regional brain volumes relative to age- and sex-matched normative data. For each anatomical region, the report displays the percentile (relative to the reference population), the volume expressed as percentage of intracranial volume (ICV), and the absolute volume in cubic centimeters. Shown metrics include: global brain volumetry (top row), hippocampal volumetry (second row), and volumetry of frontal, occipital, parietal, temporal lobes, and cerebellum (subsequent rows). Regions with percentile values below the 5th percentile are automatically highlighted in red.

matter) and subcortical (located in the white matter without reaching the cortex) and deep MBs according to their relative position to the cortex using coregistered T1-3D images. Cortical, juxtacortical, subcortical MBs were grouped into a single subgroup and divided from deep MBs (14).

Subjects with  $\geq 4$  MBs superficial were considered pathological, in accordance with literature data (15) and

FDA guidelines about the criteria for inclusion of AD Pz to anti-amyloid Ab drug therapy (16). Literature data have shown that there is an association between the presence of more than 4 microbleeds and cognitive decline. In fact, lobar MBs are associated by a decline in executive function, information processing and mnemonic function (14).

## 2.5 Statistical analysis

Differences in categorical variables between groups (patients vs. non-dementia patients) were assessed using the  $\chi^2$  test, while differences in continuous variables were evaluated with Student's *t*-test or Wilcoxon's test, depending on distribution normality assessed by the Shapiro–Wilk test (30).

The intraclass correlation coefficient (ICC, two-way random effects, absolute agreement, single measures) was used to assess inter-observer reliability in semi-quantitative ratings. Reliability was interpreted as poor ( $<0.50$ ), moderate ( $0.50$ – $0.75$ ), good ( $0.75$ – $0.90$ ), or excellent ( $>0.90$ ) (Koo & Li, 2016).

Correlations between quantitative volumetric measures from Quantib<sup>®</sup> ND and semi-quantitative atrophy scores were analyzed using Spearman's rank correlation coefficient ( $\rho$ ). The MTA score was correlated with hippocampal volume (right and left hippocampus), the Koedam score with mean parietal lobe volume, and the Kipps scale with mean frontal, temporal, and hippocampal lobe volumes. Spearman's coefficient was interpreted as negligible ( $0 < \rho < 0.20$ ), weak ( $0.21$ – $0.40$ ), moderate ( $0.41$ – $0.60$ ), strong ( $0.61$ – $0.80$ ), or very strong ( $0.81$ – $1.00$ ). Differences in correlations between readers were tested formally (17).

Cohen's Kappa was used to assess agreement in the identification of neuroradiological signs suggestive of neurodegenerative disease between semi-quantitative ratings (for both readers) and Quantib<sup>®</sup> ND software, using clinical diagnosis as the gold standard. Kappa values were interpreted according to Cohen as follows:  $0.01$ – $0.20$  (slight),  $0.21$ – $0.39$  (fair),  $0.40$ – $0.59$  (moderate),  $0.60$ – $0.79$  (substantial), and  $0.80$ – $1.00$  (almost perfect).

Diagnostic performance metrics (sensitivity, specificity, positive predictive value, negative predictive value, and accuracy) were derived from  $2 \times 2$  contingency tables. Ninety-five percent confidence intervals (95% CI) for these metrics were calculated using the Wilson score method. Differences in diagnostic accuracy among readers and Quantib<sup>®</sup> ND were tested using the method proposed by Roldán-Nofuentes and Sidaty-Regad (2019).

All thresholds applied for semi-quantitative visual scales (Scheltens, Koedam, and Kipps) were pre-specified based on established literature (references 8–10) and not derived from study data. Patients were classified as pathological or non-pathological according to these predefined cut-offs, as described in Section 2.c.

Two-sided *p*-values  $< 0.05$  were considered statistically significant. All analyses were performed using R version 4.4.1 (R Foundation for Statistical Computing, Vienna, Austria). Figure 3 graphically shows our study design and analysis pipeline.

## 3 Results

### 3.1 Demographics

143 patients were selected, 10 of whom were excluded due to lack of clinical neurological data (inconclusive diagnoses and/or

patients lost to subsequent follow-ups). Thus, a total of 133 patients were included in the study (Supplementary Table S2 shows the different diagnosis distribution).

Patients in the non-dementia group were referred to the neurology unit for subjective complaints or non-specialist clinical evaluation, but specialist clinical assessment, clinical tests, and follow-up did not result in a dementia diagnosis. Non-dementia pathological patients did not show statistically significant differences with regard to sex (*p*-value 0.73). However, differences are present with regard to age; in fact, the non-dementia patients were younger relative to the patients (mean age of 66.67 and 73.61 respectively, Mann–Whitney test *p*-value of 0.001), as shown in Table 1.

### 3.2 Concordance between expert and nonexpert observers

Observer 1 and Observer 2 showed good concordance for the MTA scale rating for both sides, for the right (ICC 0.86, 95% CI 0.81–0.90) and left (ICC 0.82, 95% CI 0.75–0.87), moderate concordance for the Koedam scale (ICC 0.66, 95% CI 0.55–0.74) and good concordance for the Kipps scale (ICC 0.76, 95% CI 0.70–0.84), respectively. Subclass analysis of Kipps scale shows that observer 1 and 2 have good concordance for frontal region (ICC 0.766 95% CI: 0.685–0.829) and posterior temporal region (ICC 0.791 95% CI: 0.706–0.852) and poor for anterior temporal region (0.589 95% CI: 0.408–0.715), as shown in Table 2.

### 3.3 Concordance between observers and correlation with volumetry

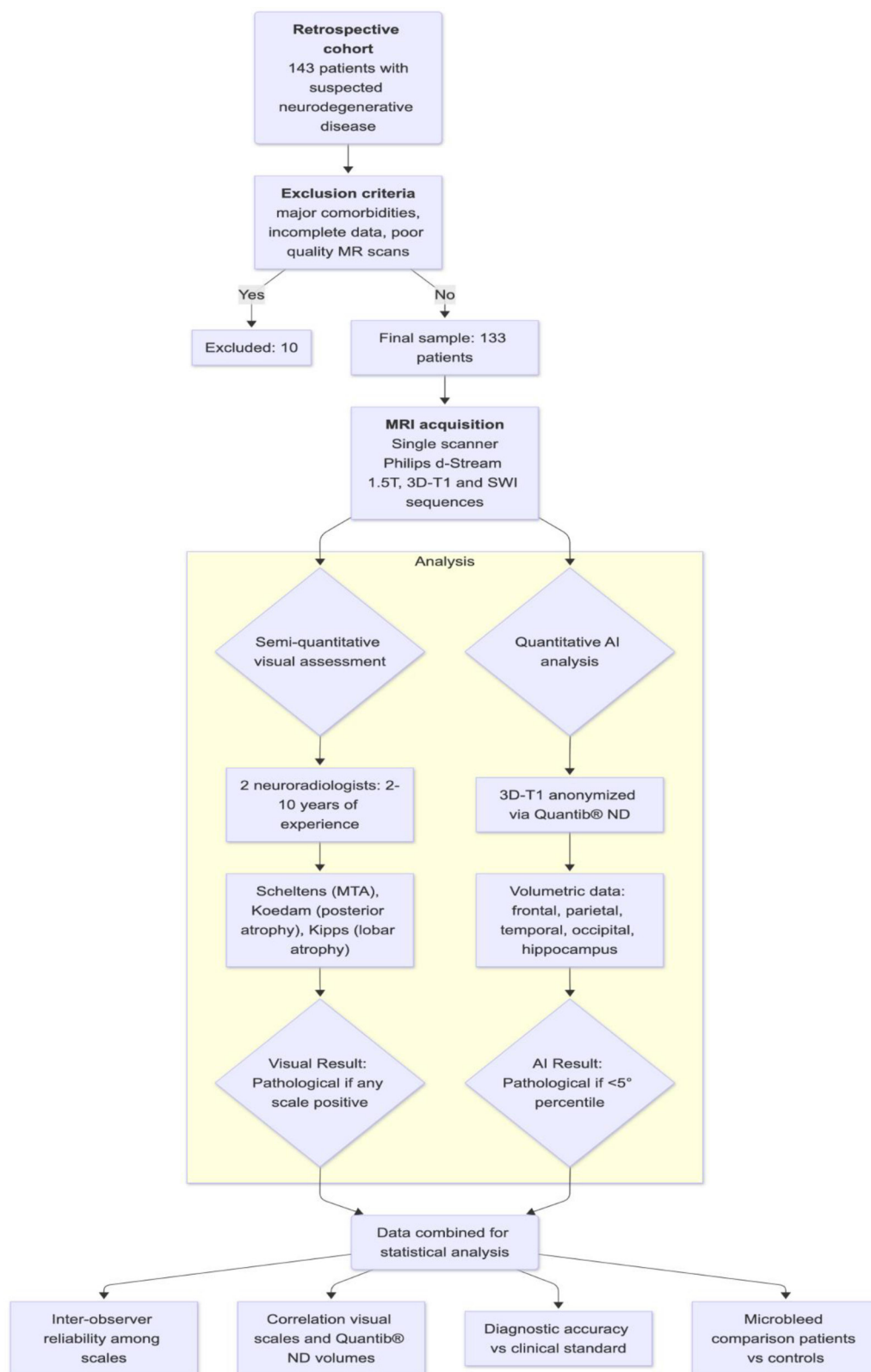
Spearman correlation values between semi-quantitative assessment and brain volumetry are shown in Table 3.

### 3.4 Concordance between observers and clinical gold-standards

Concordance between evaluators and clinical gold standard and accuracy metrics obtained from  $2 \times 2$  tables are shown in Table 4.

Observer 2 obtained higher accuracy and better Cohen's K than the Quantib<sup>®</sup> ND software and the inexperienced evaluator. The differences in Cohen's K were not statistically significant because of the overlapping 95% CI between the three groups. The difference in diagnostic accuracy of inexperienced vs. experienced reader is not statistically significant (*p*-value  $> 0.05$ ). The difference in diagnostic accuracy of Quantib<sup>®</sup> ND vs. observer 1 and 2 are statistically significant (*p*-value  $< 0.001$ ).

The assessment performed independently by the two observers with semi-quantitative scales showed greater sensitivity than the Quantib<sup>®</sup> ND assessment, which showed greater specificity.



**FIGURE 3** Schematic representation of the study workflow. A total of 133 subjects underwent 1.5 T MRI, including 3D-T1 and SWI sequences. Semi-quantitative visual assessment was performed independently by two neuroradiologists using Scheltens, Koedam, and Kipps scales. Quantitative volumetric analysis was obtained using Quantib® ND software. Microbleeds were classified on SWI images. Statistical analysis included ICC for inter-observer reliability, Spearman's correlation between visual and quantitative measures, Cohen's kappa for diagnostic agreement, and  $\chi^2$  test for microbleed frequency.

### 3.5 Differences in microbleeds

Microbleeds were assessed on SWI and classified according to their anatomical location as lobar (cortical, juxtacortical, or subcortical) or deep (basal ganglia, thalamus, brainstem, or cerebellum). In the final dataset ( $n = 137$ ), microbleed prevalence differed substantially between diagnostic groups ( $\chi^2$   $p$ -value 0.04). Among patients with dementia ( $n = 98$ ), 44 subjects (44.9%) exhibited at least one lobar microbleed, whereas this was observed in only 2 out of 39 non-dementia patients (5.1%). Deep microbleeds were also more frequent in the dementia group (17/98; 17.3%) than in the non-dementia patients group (2/39; 5.1%). Regarding burden,  $\geq 4$  lobar microbleeds were identified in 15 dementia patients (15.3%) and in only one non-dementia subject (2.6%).

TABLE 1 Sex and age characteristics of the study cohort.

Demographics	Non-dementia patients	Patients	$p$ -value
N	35	98	
Sex (M/F)	16/19	50/48	0.73
Age (years)	66.67 $\pm$ 11.18	73.61 $\pm$ 8.54	0.002

TABLE 2 Concordance between expert and non-expert observer in the evaluation of the scales used.

Visual scale	Right	Left
MTA	ICC 0.82, 95% CI 0.75–0.87),	(ICC 0.86, 95% CI 0.81–0.90)
Koedam	ICC 0.66, 95% CI 0.55–0.74	
Kipps	ICC 0.76, 95% CI 0.70–0.84	
Frontal Kipps	ICC 0.766 95% CI: 0.685–0.829	
Rear temporal kipps	ICC 0.791 95% CI: 0.706–0.852	
Front temporal kipps	ICC 0.589 95% CI: 0.408–0.715	

TABLE 3 Concordance between semi-quantitative assessment and respective brain volumetry.

Visual scale vs. cerebral region	Observer 1 ( $\rho$ , $p$ -value)	Observer 2 ( $\rho$ , $p$ -value)
MTA vs. hippocampus (right)	-0.51, <0.001	-0.60, <0.001
MTA vs. hippocampus (left)	-0.52, <0.001	-0.60, <0.001
Koedam vs. parietal lobe	-0.44, <0.001	-0.46, <0.001
Kipps vs. frontal lobe	-0.38, <0.001	-0.39, 0.007
Kipps vs. temporal lobe	-0.66, <0.001	-0.63, <0.001
Kipps vs. hippocampus	-0.53, <0.001	-0.55, <0.001

TABLE 4 Concordance between observer 1, observer 2 and quantib<sup>®</sup> ND and dementia diagnosis.

Observer/software	Cohen’s K (95% CI)	Accuracy	SE	SP	PPV	NPV
Observer 1	0.28 (0.1–0.45)	0.71 (0.62, 0.78)	0.77 (0.68, 0.85)	0.51 (0.34, 0.69)	0.82 (0.72, 0.89)	0.45 (0.29, 0.62)
Observer 2	0.39 (0.21–0.56)	0.75 (0.67, 0.82)	0.79 (0.70, 0.87)	0.62 (0.44, 0.78)	0.86 (0.77, 0.92)	0.51 (0.35, 0.67)
Quantib <sup>®</sup> ND	0.23 (0.08–0.37)	0.61 (0.52, 0.69)	0.56 (0.46, 0.66)	0.74 (0.56, 0.87)	0.86 (0.75, 0.93)	0.37 (0.26, 0.50)

Sensitivity (Se); Specificity (Sp); positive predictive value (PPV); negative predictive value (NPV).

## 4 Discussion

Magnetic resonance imaging (MRI) has long been an essential tool in diagnosing suspected dementia, not only for ruling out secondary causes but also for targeting clinical diagnoses by documenting specific brain regions and patterns of atrophy. Unlike many studies that evaluate dementia patients based on solid clinical suspicion in a research setting (6, 18–20), our study mirrors everyday clinical practice, where patients often present without a precise clinical framework. This is in addition to the absence of laboratory data or nuclear medicine imaging to corroborate the neuroradiological findings. In this context, we demonstrate how both semi-quantitative rating scales and brain volumetry processed by artificial intelligence (AI) software consistently contribute to the diagnostic process for patients with suspected dementia. Our analysis benefits from a homogeneous dataset, as all patients underwent the same MRI protocol on the same 1.5 Tesla scanner—considered the gold standard in clinical practice and the most widely used MRI magnet model in the region. This robust design strengthens our findings, ensuring consistency across the data. The first analysis compared the results of two evaluators with varying expertise in using semi-quantitative rating scales (11–13). We employed three distinct scales (Medial Temporal Lobe - MTA, Koedam, and Kipps) to evaluate different neurodegenerative disorders (Alzheimer’s, Posterior Cortical Atrophy -PCA, and Frontotemporal dementia -FTD), a necessity since our patients lacked major diagnostic suspicion or specific biomarkers. The standard neuroradiological approach was followed to assess atrophy location and severity. The final diagnosis, made through comprehensive clinical, laboratory, and instrumental evaluation, served as the gold standard.

Our results align with existing literature, showing that both raters had a strong agreement in semi-quantitative assessments (4, 18). Notably, the highest agreement was observed using the Kipps scale, followed by the Schelten scale. This may be because these scales more directly capture regions vulnerable to dementia pathology, where changes are more readily identifiable. When examining the subdomains of the Kipps scale, we found good concordance in the frontal and posterior temporal regions, but lower agreement in the anterior temporal region. This discrepancy likely arises from the anatomical challenges in assessing the anterior temporal area, which lacks clear landmarks compared to the more easily identifiable posterior temporal region. In our memory-clinic cohort, the majority of patients had a clinical diagnosis of Alzheimer’s disease. The Schelten MTA and Kipps lobar atrophy scales predominantly

target medial temporal and frontal/temporal regions, which are commonly and relatively early involved across these disorders, particularly in typical late-onset Alzheimer's disease and behavioral-variant frontotemporal dementia. This may, at least in part, explain the higher agreement we observed for these scales compared with the Koedam posterior atrophy scale. Nevertheless, we acknowledge that different dementia subtypes exhibit distinct atrophy patterns, and our mixed cohort does not allow a subtype-specific validation of each scale. We have therefore rephrased our conclusions to emphasize that our findings relate to a heterogeneous, real-world memory-clinic population rather than to any single nosological entity.

Regarding the Koedam scale, we found moderate concordance, likely due to its requirement to assess multiple cortical areas across several planes, which may complicate the agreement between raters on subtle atrophic changes (21). Additionally, the less experienced observer tended to overestimate findings, a pattern that underscores the importance of training in using these scales in clinical practice.

In our second analysis, we found moderate diagnostic accuracy between the scores generated by the two observers and the brain volumetry data processed by Quantib<sup>®</sup> ND. This validates the role of semi-quantitative scales in the diagnostic approach, as they help define site-specific atrophy and gauge severity, aligning with the findings of several published studies (4, 18, 20, 22). Interestingly, correlation between semi-quantitative ratings and volumetry was strongest for the hippocampal and posterior temporal regions, as seen with the MTA and Kipps scales. In contrast, correlation was weaker for the parietal and frontal lobes, where the Koedam and Kipps frontal subdomains showed lower agreement. This aligns with literature suggesting that Alzheimer's disease, especially in late-onset forms, primarily affects the temporo-mesial regions (23).

The third analysis examined how well semi-quantitative assessments and brain volumetry software could correlate with and predict a clinical dementia diagnosis. Our findings highlighted a difference between the more experienced and less experienced observers, emphasizing the need for adequate training in using these visual scales (4, 6, 21). Notably, visual rating scales demonstrated greater diagnostic sensitivity than brain volumetry alone, though the latter showed higher specificity. This suggests that volumetric data should not be relied upon in isolation; rather, it must be interpreted alongside semi-quantitative grading. The greater sensitivity of the semi-quantitative scales aids in capturing a broader range of dementia cases, albeit at the cost of increased false positives, whereas brain volumetry's specificity helps exclude definitively non-pathological cases.

These results highlight the potential synergy between AI-based brain volumetry and visual rating scales in diagnosing cognitive impairment. Specifically, using volumetric data to rule out non-pathological cases could prevent unnecessary follow-up exams, reduce healthcare costs, and streamline patient management. Previous studies have shown that combining AI and semi-quantitative scales improves diagnostic accuracy and reduces false positives, especially for Alzheimer's and FTD patients, and

enhances the diagnostic performance of less experienced readers (18, 23, 24). While some studies question the clinical utility of AI in this context (25–27), we believe the contradictory results may stem from inconsistent normative data and non-standardized imaging protocols. To address this, our study used a consistent MRI protocol with FDA and CE-approved software, ensuring robust, reproducible results.

Our final analysis examined the presence of microbleeds (MBs), revealing a correlation between lobar MBs and a clinical dementia diagnosis. Lobar microbleeds (MBs) play an increasingly important role in the differential diagnosis and management of patients with cognitive impairment. Their presence can help distinguish neurodegenerative disease-related atrophy from vascular or amyloid angiopathy-related pathology, particularly when MBs are predominantly cortical or juxtacortical. This distinction is clinically relevant, as extensive lobar MBs are often associated with cerebral amyloid angiopathy and may contraindicate or complicate anti-amyloid monoclonal antibody therapy due to the risk of amyloid-related imaging abnormalities (ARIA-A and ARIA-E). Consistent with previous studies (28), our findings support the inclusion of MB assessment as a complementary imaging biomarker in the diagnostic work-up of dementia, contributing both to etiologic differentiation and to safer therapeutic decision-making (16, 18).

Although AI-based volumetric tools such as Quantib<sup>®</sup> ND currently serve as complementary aids rather than replacements for expert visual assessment, they hold considerable potential to improve diagnostic workflow in the near future. By providing standardized, automated quantification of regional brain atrophy, these systems can significantly reduce reporting time, increase consistency across readers, and support faster clinical decision-making, particularly in high-volume or general radiology settings. In addition, the integration of such tools into routine practice may streamline diagnostic pathways and ultimately lower healthcare costs by reducing the need for repeated or ancillary examinations. Future cost-effectiveness and workflow studies are warranted to confirm these anticipated advantages and to define the optimal balance between automated analysis and expert interpretation. Looking forward, it would be valuable to replicate this analysis in a similar cohort using a 3 T MRI scanner and conduct follow-up MRIs to track how quantitative and semi-quantitative data evolve over time, potentially identifying early radiological signs of dementia.

A key limitation of our study is the absence of a community-based healthy control group; instead, our comparator group consisted of patients referred to the memory clinic with cognitive complaints or subjective memory concerns in whom a dementia diagnosis was ultimately ruled out after specialist assessment and follow-up. Also, the exclusion of certain semi-quantitative scales like GCA (29) or those more sensitive, such as those described by Harper et al. (4). We also did not correlate scale scores with specific dementia patterns but rather with a broad clinical diagnosis of dementia. The proportion of dementia vs. non-dementia cases reflects the case-mix of a tertiary memory clinic and resulted in unbalanced group sizes. Although this does not invalidate the calculation of diagnostic

performance metrics derived from  $2 \times 2$  contingency tables, it may limit generalizability to settings with a different prevalence of dementia. Additionally, we did not conduct follow-up MRIs, which could confirm or refine our initial findings and track the progression of atrophy. Lastly, the Kipps scale, which separates the anterior and posterior temporal lobes, may not fully align with volumetric analyses that assess the entire temporal lobe, introducing potential discrepancies in the findings. Microbleed assessment was performed by a single neuroradiologist, which prevented evaluation of inter-rater reliability and may have introduced observer-related bias in MB detection. Although our findings demonstrate good agreement between imaging-based evaluations and clinical diagnosis, a key limitation of this study is the absence of correlation with neuropsychological performance, disease severity scales, or biological markers such as cerebrospinal fluid (CSF) and positron emission tomography (PET) data. These parameters could have provided additional validation of the imaging results and strengthened the clinical relevance of our conclusions. Future prospective studies combining MRI with biomarker and cognitive assessments are warranted to establish a more comprehensive diagnostic model that integrates structural, functional, and biochemical measures of neurodegeneration.

## 5 Conclusions

Our study thus validated the role of semi-quantitative rating scales as an effective and sensitive tool in the neuroradiological diagnostic approach to the patient with suspected dementia. The use of artificial intelligence software to quantify the brain atrophy finding is also an important supplementary element to the diagnostic work-up as it ensures greater homogeneity of results when used by observers with different degrees of experience, especially allowing the exclusion of pictures that are not definitely pathological to the imaging finding.

## Data availability statement

The datasets presented in this article are not readily available. Requests to access the datasets should be directed to calloni.sonia@hsr.it.

## Ethics statement

The studies involving humans were approved by San Raffaele Hospital Ethic Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

SC: Writing – original draft, Investigation, Conceptualization, Writing – review & editing. AD: Writing – review & editing, Conceptualization, Writing – original draft, Investigation. GA: Writing – review & editing, Writing – original draft, Formal analysis, Investigation, Data curation, Conceptualization, Methodology. MZ: Formal analysis, Data curation, Writing – review & editing, Methodology. PV: Conceptualization, Supervision, Investigation, Writing – review & editing. GC: Writing – review & editing, Visualization, Investigation, Supervision. ES: Writing – review & editing, Investigation, Formal analysis, Data curation. GR: Formal analysis, Data curation, Writing – review & editing, Conceptualization, Investigation. AG: Data curation, Writing – review & editing, Methodology, Conceptualization, Formal analysis. GM: Validation, Writing – review & editing, Resources, Visualization. FC: Writing – review & editing, Data curation, Methodology, Validation, Visualization. AL: Supervision, Software, Writing – review & editing. FA: Visualization, Validation, Writing – review & editing, Supervision. MF: Visualization, Validation, Supervision, Writing – review & editing. AF: Visualization, Conceptualization, Validation, Project administration, Supervision, Writing – review & editing.

## Funding

The author(s) declared that financial support was not received for this work and/or its publication.

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent

those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Mitchell AJ, Beaumont H, Ferguson D, Yadegarfar M, Stubbs B. Risk of dementia and mild cognitive impairment in older people with subjective memory complaints: meta-analysis. *Acta Psychiatr Scand.* (2014) 130(6):439–51. doi: 10.1111/acps.12336
- Rizzi L, Rosset I, Roriz-Cruz M. Global epidemiology of dementia: alzheimer's and vascular types. *BioMed Res Int.* (2014) 2014:1–8. doi: 10.1155/2014/908915
- Frisoni GB, Fox NC, Jack CR, Scheltens P, Thompson PM. The clinical use of structural MRI in Alzheimer's disease. *Nat Rev Neurol.* (2010) 6(2):67–77. doi: 10.1038/nrneuro.2009.215
- Harper L, Fumagalli GG, Barkhof F, Scheltens P, O'Brien JT, Bouwman F, et al. MRI Visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. *Brain.* (2016) 139(4):1211–25. doi: 10.1093/brain/aww005
- Giorgio A, De Stefano N. Clinical use of brain volumetry. *J Magn Reson Imaging.* (2013) 37(1):1–14. doi: 10.1002/jmri.23671
- Pemberton HG, Goodkin O, Prados F, Das RK, Vos SB, Moggridge J, et al. Automated quantitative MRI volumetry reports support diagnostic interpretation in dementia: a multi-rater, clinical accuracy study. *Eur Radiol.* (2021) 31:5312–23. doi: 10.1007/s00330-020-07455-8
- Zhang Y, Wang X, Chen J, et al. Artificial intelligence-driven MRI analysis for early detection and classification of dementia: a comprehensive review. *IEEE Transactions on Cybernetics.* (2023) 53(12).
- Li H, Zhao Q, Sun Y, et al. Deep learning and explainable artificial intelligence in neuroimaging of Alzheimer's disease: current advances and future perspectives. *IEEE Transact Pattern Anal Mach Intell.* (2024) 46(4).
- Pemberton HG, Zaki LAM, Goodkin O, Das RK, Steketee RME, Barkhof F, et al. Technical and clinical validation of commercial automated volumetric MRI tools for dementia diagnosis—a systematic review. *Neuroradiology.* (2021) 63(11):1773–89. doi: 10.1007/s00234-021-02746-3
- Sperling RA, Jack CR Jr, Black SE, Frosch MP, Greenberg SM, Hyman BT, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: recommendations from the Alzheimer's association research roundtable workgroup. *Alzheimers Dement.* (2011) 7(4):367–85. doi: 10.1016/j.jalz.2011.05.2351
- Velickaitė V, Ferreira D, Cavallin L, Lind L, Ahlström H, Kilander L, et al. Medial temporal lobe atrophy ratings in a large 75-year-old population-based cohort: gender-corrected and education-corrected normative data. *Eur Radiol.* (2018) 28(4):1739–47. doi: 10.1007/s00330-017-5103-6
- Kipps CM, Davies RR, Mitchell J, Kril JJ, Halliday GM, Hodges JR. Clinical significance of lobar atrophy in frontotemporal dementia: application of an MRI visual rating scale. *Dement Geriatr Cogn Disord.* (2007) 23(5):334–42. doi: 10.1159/000100973
- 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(12):2618–25. doi: 10.1007/s00330-011-2205-4
- Akoudad S, Wolters FJ, Viswanathan A, de Bruijn RF, van der Lugt A, Hofman A, et al. Association of cerebral microbleeds with cognitive decline and dementia. *JAMA Neurol.* (2016) 73(8):934–43. doi: 10.1001/jamaneurol.2016.1017
- Withington CG, Turner RS. Amyloid-Related imaging abnormalities with anti-amyloid antibodies for the treatment of dementia due to Alzheimer's disease. *Front Neurol.* (2022) 13:862369. doi: 10.3389/fneur.2022.862369
- Available online at: [www.accessdata.fda.gov/Drugsatfda\\_docs/Label/2023/761269Orig1s001lbl.Pdf](http://www.accessdata.fda.gov/Drugsatfda_docs/Label/2023/761269Orig1s001lbl.Pdf) (Accessed December 30, 2025).
- Hittner JB, May K, Silver NC. A monte carlo evaluation of tests for comparing dependent correlations. *J Gen Psychol.* (2003) 130(2):149–68. doi: 10.1080/00221300309601282
- Calloni SF, Vezzulli PQ, Castellano A, Leone R, Basaia S, von Loon A, et al. Combining semi-quantitative rating and automated brain volumetry in MRI evaluation of patients with probable behavioral variant of fronto-temporal dementia: an added value for clinical practise? *Neuroradiology.* (2023) 65(6):1025–35. doi: 10.1007/s00234-023-03133-w
- Bartos A, Gregus D, Ibrahim I, Tintèra J. Brain volumes and their ratios in Alzheimer's disease on magnetic resonance imaging segmented using freesurfer 6.0. *Psychiatry Res Neuroimaging.* (2019) 287:70–4. doi: 10.1016/j.psychres.2019.01.014
- Möller C, van der Flier WM, Versteeg A, Benedictus MR, Wattjes MP, Koedam EL, et al. Quantitative regional validation of the visual rating scale for posterior cortical atrophy. *Eur Radiol.* (2014) 24(2):397–404. doi: 10.1007/s00330-013-3025-5
- Rhodus-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. doi: 10.3389/fnagi.2017.00117
- Vernooij MW, Jaspers B, Steketee R, Koek M, Vrooman H, Ikram MA, et al. Automatic normative quantification of brain tissue volume to support the diagnosis of dementia: a clinical evaluation of diagnostic accuracy. *NeuroImage Clin.* (2018) 20:374–9. doi: 10.1016/j.nicl.2018.08.004
- Amland R, Selbaek G, Brækhus A, Edwin TH, Engedal K, Knapskog AB, et al. Clinically feasible automated MRI volumetry of the brain as a prognostic marker in subjective and mild cognitive impairment. *Front Neurol.* (2024) 15:1425502. doi: 10.3389/fneur.2024.1425502
- Mai Y, Cao Z, Zhao L, Yu Q, Xu J, Liu W, et al. The role of visual rating and automated brain volumetry in early detection and differential diagnosis of Alzheimer's disease. *CNS Neurosci Ther.* (2024) 30(4):e14492. doi: 10.1111/cns.14492
- Rudolph J, Rueckel J, Döpfert J, Ling WX, Opalka J, Brem C, et al. Artificial intelligence-based rapid brain volumetry substantially improves differential diagnosis in dementia. *Alzheimers Dement Diagn Assess Dis Monit.* (2024) 16(4):e70037. doi: 10.1002/dad2.70037
- Ingala S, van Maurik IS, Altomare D, Wurm R, Dicks E, van Schijndel RA, et al. Clinical applicability of quantitative atrophy measures on MRI in patients suspected of Alzheimer's disease. *Eur Radiol.* (2022) 32(11):7789–99. doi: 10.1007/s00330-021-08503-7
- Zilioli A, Rosenberg A, Mohanty R, Matton A, Granberg T, Hagman G, et al. Brain MRI volumetry and atrophy rating scales as predictors of amyloid status and eligibility for anti-amyloid treatment in a real-world memory clinic setting. *J Neurol.* (2024) 272(1):84. doi: 10.1007/s00415-024-12853-9
- Withington CG, Turner RS. Amyloid-Related imaging abnormalities with anti-amyloid antibodies for the treatment of dementia due to Alzheimer's Disease. *Front Neurol.* (2022) 13:862369. doi: 10.3389/fneur.2022.862369
- Pasquier F, Leys D, Weerts JG, Mounier-Vehier F, Barkhof F, Scheltens P. Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. *Eur Neurol.* (1996) 36(5):268–72. doi: 10.1159/000117270
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* (2016) 15(2):155–63. doi: 10.1016/j.jcm.2016.02.012

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fradi.2026.1667306/full#supplementary-material>