1   **Interpretable MRI-Based Deep Learning for Alzheimer's Risk and Progression**

2   Bin Lu,[1,2,3#] Yan-Rong Chen,[4,#] Rui-Xian Li,[5,#] Ming-Kai Zhang,[5] Shao-Zhen Yan,[6] Guan-Qun

3   Chen,[7] Francisco Xavier Castellanos,[8] Paul M. Thompson,[9] Jie Lu,[6,*] Ying Han,[5,10,11,12,13*] Chao-

4   Gan Yan[1,2,4,14,*]

5   **Affiliations:**

6   [1]State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese

7   Academy of Sciences, Beijing 100101, China

8   [2]Department of Psychology, University of the Chinese Academy of Sciences, Beijing 100049,

9   China

10   [3]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour,

11   Radboud University Nijmegen Medical Centre, Nijmegen, 6525 EN, the Netherlands

12   [4]Magnetic Resonance Imaging Research Center, Institute of Psychology, Chinese Academy of

13   Sciences, Beijing 100101, China

14   [5]Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing 100053,

15   China

16   [6]Department of Radiology and Nuclear Medicine, Xuanwu Hospital, Capital Medical University,

17   Beijing 100053, China

18   [7]Department of Neurology, Beijing Chaoyang Hospital, Capital Medical University, No.8 Gongti

19   South Road, Chaoyang District, Beijing 100020, China

20   [8]Department of Child and Adolescent Psychiatry, New York University Grossman School of

21   Medicine, New York, NY 10016, USA

22   [9]Imaging Genetics Center, Mark & Mary Stevens Institute for Neuroimaging & Informatics,

23   Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

24   [10]State Key Laboratory of Digital Medical Engineering, Key Laboratory of Biomedical

25   Engineering of Hainan Province, School of Biomedical Engineering, Hainan University, Sanya,

26   Hainan 572025, China

27 [11]The Institute of Neurological and Psychiatric Disorders, Shenzhen Bay Laboratory, Shenzhen,

28 Guangdong 518132, China

29 [12]National Clinical Research Center for Geriatric Diseases, Beijing 100053, China

30 [13]The Central Hospital of Karamay, Xinjiang 834000, China

31 [14]Department of Psychological and Cognitive Sciences, Tsinghua University, Beijing 100084,

32 China

33

34 # These authors contributed equally as the first author.

35 *Corresponding authors:

36 Chao-Gan Yan, Ph.D. Address: Room 304, Weiqing Building, Tsinghua University, Haidian

37 District, Beijing, 100084. E-mail: yancg@tsinghua.edu.cn Telephone: (+86) 131-4614-8848.

38 Ying Han, Ph.D. Address: 45 Changchun Street, Xicheng District, Beijing 100053, China. E-

39 mail: hanying@xwh.ccmu.edu.cn, Telephone: (+86) 136-2101-1941.

40 Jie Lu, Ph.D. Address: 45 Changchun Street, Xicheng District, Beijing 100053, China. E-mail:

41 imaginglu@hotmail.com, Telephone: (+86) 189-1150-7070.

42

43   **ABSTRACT**

44   Timely intervention for Alzheimer's disease (AD) requires early detection. The development of

45   immunotherapies targeting amyloid-beta and tau underscores the need for accessible, time-

46   efficient biomarkers for early diagnosis. Here, we directly applied our previously developed

47   MRI-based deep learning model for AD to the large Chinese SILCODE cohort (722 participants,

48   1,105 brain MRI scans). The model — initially trained on North American data — demonstrated

49   robust cross-ethnic generalization, without any retraining or fine-tuning, achieving an AUC of

50   91.3% in AD classification with a sensitivity of 95.2%. It successfully identified 86.7% of

51   individuals at risk of AD progression more than 5 years in advance. Individuals identified as

52   high-risk exhibited significantly shorter median progression times. By integrating an

53   interpretable deep learning brain risk map approach, we identified AD brain subtypes, including

54   an MCI subtype associated with rapid cognitive decline. The model's risk scores showed

55   significant correlations with cognitive measures and plasma biomarkers, such as tau proteins and

56   neurofilament light chain (NfL). These findings underscore the exceptional generalizability and

57   clinical utility of MRI-based deep learning models, especially in large and diverse populations,

58   offering valuable tools for early therapeutic intervention. The model has been made open-source

59   and deployed to a free online website for AD risk prediction, to assist in early screening and

60   intervention.

61

INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disease that poses a significant challenge to global health, profoundly impacting individuals, families, and healthcare systems[1]. Early detection of AD is crucial, as it allows for timely interventions that could slow disease progression and improve patient outcomes[1]. The US Food & Drug Administration recently granted traditional approval for lecanemab and donanemab for treating mild cognitive impairment (MCI) and mild AD in patients with amyloid-beta (Aβ) proteinopathy or pathologic tau deposition[2,3]. The advent of these novel immunotherapies highlights the need for cost-effective and time-efficient biomarkers for early diagnosis to enhance treatment effectiveness[4]. Traditionally, AD diagnosis has relied on observing cognitive decline, which typically occurs in later stages, often missing the critical window for effective intervention. Since biological changes often precede cognitive symptoms, integrating biomarkers from cerebrospinal fluid (CSF) and positron emission tomography (PET) into AD diagnostic guidelines has significantly enhanced early detection[5]. However, the invasive nature, high costs and need for specially trained personnel to implement these methods restrict their widespread clinical application.

Alternatively, magnetic resonance imaging (MRI) could provide a non-invasive and more accessible substitute. Currently, structural brain MRI is primarily used to monitor the dynamic progression of AD pathology, including staging, predicting prognosis, and patient selection and safety monitoring during drug treatment[6,7]. However, the advent of deep learning methods has revolutionized MRI's potential by enabling the detection and use of subtle features previously unrecognizable by human experts, thus advancing the field towards earlier and more precise diagnostics and facilitating timely medical treatment[8–10]. The development of interpretable

84   models has opened the "black box" of deep learning, further enhancing the confidence of

85   physicians and patients in medical imaging models applied to clinical practice[11–13].

86   Despite enthusiasm for deep learning in AD diagnosis, significant limitations exist. First,

87   robust independent validation is often lacking. Recent systematic reviews of MRI-based AD

88   classifiers revealed startling statistics: only 0–5.3% of studies (e.g., 2 of 81[14]; 0 of 16[15]; and 6 of

89   114[16]) used any independent validation dataset, never mind including cross-ethnic validation.

90   Second, many classifiers have been trained on relatively small sample sizes, leading to models

91   that are susceptible to overfitting. These two factors have contributed to poor generalizability,

92   rendering many models impractical for clinical application. In response to these challenges, we

93   employed transfer learning to construct a generalizable AD classification model based on large

94   and diverse structural MRI datasets, encompassing over 85,000 samples from 217 different

95   sites/scanners[17]. This model achieved an accuracy rate exceeding 90% across four independent

96   AD datasets. Nevertheless, the prior study identified several areas requiring further work: first,

97   the training and validation datasets primarily comprised Western samples, raising concerns about

98   generalizability to other populations, such as those in China; second, the model's capability to

99   predict disease progression needed further development, as this is essential for selecting patients

100   with early-stage AD for timely pharmacological treatment; third, the pathological relevance of an

101   imaging-based model should be validated by comparison with established biomarkers and

102   cognitive measurements; finally, providing individualized AD-risk brain maps can enhance the

103   model's interpretability and facilitate monitoring treatment efficacy. Recently, plasma-based

104   biomarkers have shown both feasibility and high accuracy in detecting AD pathology and

105   predicting progression of cognitive decline[18,19]. In the 2024 revision of the NIA-AA diagnostic

106   criteria, plasma analytes such as phosphorylated tau 217 immunoassay were designated as

107   belonging to the "core-1" category, equivalent to CSF and Amyloid PET[20]. Cross-validating

108   these two types of clinical feasibility biomarkers (plasma-based and MRI-based metrics) holds

109   significant practical implications.

110   To achieve these objectives, we applied an MRI-based classification model to a non-

111   Western, longitudinal, richly-phenotyped cohort, SILCODE, which consists of a large elderly

112   Chinese population, assessed across the stages of the AD continuum[21]. This cohort includes

113   extensive cognitive assessments, plasma biomarkers, and brain images. In the current study, we

114   validated the MRI-based AD model based on its classification accuracy, predicting patients'

115   disease progression, and its association with cognitive and pathological markers using the

116   SILCODE sample. Additionally, we developed an interpretable method to enhance its clinical

117   credibility.

118   **MATERIALS AND METHODS**

119   **Participants and procedures**

120   We included participants from the Sino Longitudinal Study on Cognitive Decline

121   (SILCODE), a registered ongoing longitudinal clinical trial (ClinicalTrials.gov identifier:

122   NCT02225964) focused on AD in the elderly Han population in mainland China. This study was

123   approved by the ethics committee of Xuanwu Hospital of Capital Medical University, and all

124   participants or their guardians provided written informed consent before enrollment. The detailed

125   SILCODE protocol has been previously described[22]. Participants were continuously recruited

126   and followed up with flexible visit intervals, initially set at 15 months and adjusted based on

127   cognitive evaluations administered by phone at three-month intervals, leading to varying follow-

128   up times and visit intervals for participants. During each visit, participants were invited to

129 Xuanwu Hospital for clinical evaluation, neuropsychological assessments, blood tests, and MRI

130 scans.

131 The current study initially included 738 participants, all over age 50 years at the time of

132 their first MRI scan, who had completed at least one MRI session. After quality control, 16

133 participants were excluded due to invalid MRI scans, leaving a final sample of 722 participants

134 (60.8% female; mean age = 66.1 ± 7.7 years at baseline). Diagnostic distribution at baseline was:

135 211 normal cognition (NC), 235 subjective cognitive decline (SCD), 193 MCI, and 83 AD.

136 Participants and their visit sessions were further filtered based on the requirements of specific

137 analyses.

Diagnostic criteria

139 During each visit, participants were classified into AD, MCI, SCD, and NC groups, with re-

140 evaluations in subsequent visits to detect changes[22]. AD diagnosis followed NIA-AA

141 guidelines[23], supplemented by a global Clinical Dementia Rating[24] (CDR) score of at least 1.

142 MCI was diagnosed using the Jak/Bondi method[25] with neuropsychological tests (see below) and

143 thresholds adjusted for the Chinese population[22]. SCD classification followed Jessen et al.'s

144 criteria[26] for SCD plus. NC was assigned to those without subjective or informant-reported

145 memory decline or evidence of cognitive impairment. In overlapping cases, the more severe

146 condition was designated. All clinical diagnoses were made by qualified physicians.

Neuropsychological tests

148 A battery of neuropsychological tests was used to measure cognitive levels of participants,

149 focusing on both global domain and specific subdomains including memory, language, and

150 executive functions. Global cognitive function was assessed using the Mini-Mental State

151 Examination[27] (MMSE), Memory and Executive Screening[28] (MES), and the Montreal Cognitive

152 Assessment Basic[29] (MOCA-B). For specific subdomains, memory tests included the Auditory

153 Verbal Learning Test—Huashan version[30]—Long delayed memory (AVLT-L) and Recognition

154 (AVLT-R); the language tests comprised the Animal Fluency Test[31] (AFT) and 30-item Boston

155 Naming Test[32] (BNT); and executive functions were evaluated using the Shape Trail Test[33] Parts

156 A (STT-A) and B (STT-B). All cognitive assessments were administered by trained

157 neuropsychologists.

158 **Biological tests**

159 Plasma levels of Aβ42, Aβ40, NfL, GFAP, p-tau 217, and p-tau 181 were measured by

160 immunoassay using a Simoa HD-X analyzer (Quanterix, Billerica, MA, USA), as per the

161 manufacturer's protocols. Specifically, plasma samples were thawed at room temperature and

162 then centrifuged for 5 minutes at 4°C and 10,000 × g. The resultant supernatant was transferred

163 to 96-well plates and diluted four-fold with sample diluent, followed by a two-step digital

164 immunoassay. The concentrations of Aβ42, Aβ40, NfL, and GFAP were measured with the

165 Simoa Neurology 4-Plex E Advantage Kit (Quanterix, Billerica, MA, USA; Cat# 103670). The

166 ALZpath Simoa pTau-217 v2 Assay Kit (Quanterix, Cat# 104371) was used to measure p-tau

167 217, and the Simoa pTau-181 Advantage V2 Kit (Quanterix, Cat# 103714) was used to measure

168 p-tau 181. Assays were performed using kits from the same lot, with all samples run in duplicate.

169 Seven or eight calibrators (2x) and two quality controls (2x) were run on each plate for each

170 analyte. The lower limits of detection were 0.136 pg/mL for Aβ42, 0.384 pg/mL for Aβ40, 0.090

171 pg/mL for NfL, 0.441 pg/mL for GFAP, 0.019 pg/mL for p-tau 217, and 0.028 pg/mL for p-tau

172 181. The lower limits of quantification were 0.378 pg/mL for Aβ42, 1.02 pg/mL for Aβ40, 0.400

173 pg/mL for NfL, 2.89 pg/mL for GFAP, 0.061 pg/mL for p-tau 217, and 0.338 pg/mL for p-tau

174   181. The intra-assay coefficients of variation for controls ranged from 2% to 13% for Aβ42, 1%

175   to 5% for Aβ40, 2% to 12% for NfL, 1% to 8% for GFAP, 2% to 11% for p-tau 217, and 1% to

176   10% for p-tau 181.

177   **MRI acquisition, preprocessing, and AD risk score generation**

178   T1-weighted structural brain MRI data were collected using a single 3.0 T GE SIGNA^{TM}

179   PET/MR scanner (GE Healthcare, Milwaukee, WI, USA) at Xuanwu Hospital of Capital

180   Medical University. Scanning parameters were as follows: SPGR sequence with inversion

181   preparation, repetition time (TR)=6.9 ms, echo time (TE)=2.98 ms, inversion time (TI)=450 ms,

182   slice number=192, slice thickness=1 mm, gap=0, flip angle=12°, field of view (FOV)=256×256

183   mm^2, matrix=256×256, voxel size=1×1×1 mm^3.

184   Manual quality control was conducted to exclude low-quality images with notable artifacts

185   or incomplete brain coverage. Images were then preprocessed using the sMRI preprocessing

186   pipeline from the Data Processing Assistant for Resting-State fMRI (DPARSF) Toolbox. GMD

187   and GMV maps were generated using the Voxel-Based Morphometry (VBM) module in the

188   toolbox. These maps were then directly input to the deep learning AD classifier to compute AD

189   risk scores. More information on preprocessing, feature extraction, and risk score generation

190   processes is available in the previous study[34] on the AD classifier.

191   In this study, two types of risk scores were computed: sigmoid and linear. Sigmoid risk

192   scores were obtained using sigmoid activation in the deep learning model, as per the previous

193   AD classifier paper[34]. Linear risk scores, produced using linear activation, were additionally

194   calculated due to their distribution being more suitable for statistical testing, in that they are

195   closer to the normal distribution than sigmoid scores. Accordingly, we used sigmoid scores for

196   classification-related analysis, and linear scores for statistical tests such as linear regression.

Data analysis

We first evaluated the performance of the MRI-based AD classifier on the Chinese
SILCODE cohort using receiver operating characteristic (ROC) curve analysis. The area under
the curve (AUC) was calculated, with 95% confidence intervals (CIs) estimated using 5,000
bootstrap samples. The optimal cutoff score was determined using the Youden index[35]. At this
threshold, we calculated accuracy, sensitivity and specificity to assess classifier performance.

Next, we applied a linear mixed-effects model to examine the association between classifier
linear scores and clinical diagnoses (NC, SCD, MCI, AD), controlling for sex, age, and
education. Participants were included as a random effect to account for the non-independence of
repeated measures, as some participants had multiple visits. Post-hoc pairwise comparisons with
Tukey correction were performed to examine differences between diagnostic groups. The
Satterthwaite approximation was used to adjust degrees of freedom for fixed effects, accounting
for the uncertainty introduced by random factors, to provide more robust results. Post-hoc
pairwise comparisons were conducted using estimated marginal means with Tukey correction to
examine differences between diagnostic groups.

Survival analyses were conducted to examine the probability of progression to AD in NC,
SCD and MCI participants. Participants were divided into high-risk and low-risk groups based
on AD classifier risk scores at baseline, with the cutoff (i.e., 0.38) identified through ROC
analysis using the Youden index. Kaplan-Meier survival curves were generated to illustrate the
probability of participants maintaining non-AD status over the follow-up period, with the log-
rank test comparing survival distributions between groups.

Additionally, visit sessions from all AD converters were analyzed to assess how early the
classifier could predict conversion. The proportion of high-risk sessions (score ≥ 0.38) at

220 different time intervals before the first AD diagnosis was used to calculate the capture rate. Mean

221 risk scores were also computed to observe trends in AD risk over time. Due to the limited data,

222 particularly the decrease in data size as the time before diagnosis increased, we categorized time

223 intervals to simplify analyses. Intervals were set as within 1 year, 1–2 years, 2–3 years, 3–5

224 years, and 5–14 years before the first AD diagnosis (**Figure S1**).

225   To examine the associations of MRI-based risk scores with cognitive tests and plasma

226 biomarkers, separate models were fitted for each measure. Cognitive and plasma outliers

227 exceeding six times the interquartile range (6IQR)[36] were excluded to minimize their influence.

228 Global associations were first assessed across the entire sample to identify general patterns,

229 followed by diagnosis-specific analyses within pre-AD phases (NC, SCD, MCI) to explore

230 phase-specific associations critical for early detection and intervention. Linear mixed-effects

231 models were generally used, with the linear risk score as the independent variable, sex, age, and

232 education as covariates, and participant included as a random effect to account for repeated

233 measures. However, for plasma biomarker models within the MCI group, general linear models

234 without the random effect were applied due to the limited number of repeated measures per

235 participant (e.g., 34 participants had only 37 plasma p-tau 217 measures), which caused

236 convergence issues in mixed-effects models. Effect sizes were measured using Cohen's f^2, which

237 reflects the extent to which the inclusion of a specific predictor variable improves the model's

238 explanatory power. The Satterthwaite approximation was used to adjust degrees of freedom for

239 fixed effects in mixed-effects models. Post-hoc pairwise comparisons were performed using the

240 estimated marginal means with Tukey correction in all models.

241   Differences across brain subtypes (see the next section for details on identification) in

242 variables of interest, such as MMSE decline rate, baseline MMSE level, baseline linear risk

243   score, and demographic variables, were assessed using ANOVA, followed by post-hoc pairwise

244   comparisons of estimated marginal means with LSD correction.

245   All analyses and visualizations were conducted using R 4.4.0[37], along with the following

246   packages: pROC[38] (for ROC analysis), lme4[39] and lmerTest[40] (for linear mixed models),

247   survival[41] and survminer[42] (for survival analysis), and emmeans[43] (for post-hoc comparisons).

248   **Identifying high-risk brain regions associated with AD**

249   We developed an interpretable approach to identify brain regions at risk for AD on an

250   individual level. An AD brain risk map can be generated for each NC, SCD, MCI, or AD

251   participant, in which brain regions with higher values are more closely associated with AD. The

252   detailed calculation process is as follows.

253   Initially, we created a simulated patient whom the classifier identified as being in an exact

254   intermediate state between AD and NC. To achieve this goal, we conducted 10,000 random

255   samplings within the AD and NC samples of the ADNI 1, ADNI GO and ADNI 2 datasets[44]. For

256   each sampling, we randomly selected 100 samples (e.g., 48 samples from the AD group and 52

257   samples from the NC group) and averaged their GMD and GMV maps to generate an average

258   template. Subsequently, the MRI-based AD classifier[34] was used to predict AD risk scores for

259   10,000 simulated patients. The GMD and GMV maps of the simulated patient whose risk score

260   was closest to 0.5 were selected as the final brain risk map template (the AD risk score for the

261   final template was 0.499928).

262   To predict the AD risk regions for a single participant, we sequentially replaced region by

263   region in the aforementioned simulated template with the corresponding real data from the

264   participant's brain. Then, we used the MRI-based AD classifier to predict the AD risk score of

265   the reconstituted brain. For instance, when the left hippocampal-related brain region of an AD

266   patient was replaced onto the simulated template, the predicted AD risk score rose from 0.5 to

267   0.88, indicating that the left hippocampus's area risk level for that patient was 0.88. Repeating

268   this process and then reprojecting the risk scores of every brain area back onto the brain

269   generates the subject's AD risk distribution map. Each replaced brain region was defined as a

270   cube with an edge length of 18 mm, and the step length for moving and replacing brain regions

271   was 12 mm.

272   Subsequently, we employed k-means cluster analysis to examine different AD risk brain

273   region patterns within the AD patients in the SILCODE dataset. We determined the optimal

274   number of clusters for k-means clustering by employing the elbow method, which utilizes the

275   sum of squared errors (SSE) to identify the point at which the increase in the number of clusters

276   does not result in significantly better modeling of the data. The center of the feature space for

277   each cluster is considered to be a representative pattern of AD-related brain abnormalities, or

278   alternatively, a demonstration of an AD subtype based on MRI characteristics.

279   **RESULTS**

280   **Study participants**

281   The current study included 722 participants from the SILCODE dataset (60.8% female; 66.1

282   ± 7.7 years of age at baseline), categorized based on clinical assessments at baseline into 211

283   NC, 235 SCD, 193 MCI, and 83 AD participants. They completed 1,980 visits and provided

284   1,105 valid MRI scans. **Table 1** presents descriptive data for these MRI sessions (60.8% female;

285   66.6 ± 7.6 years; 348 NC, 363 SCD, 269 MCI, 125 AD), which serve as the basis for the

286   following validations of classificatory ability and pathological value of the MRI-based deep

287   learning model.

288   **Table 1. Descriptive Data for Visit Sessions with Valid MRI Scans.**

	ALL (*n* = 1105)	NC (*n* = 348)	SCD (*n* = 363)	MCI (*n* = 269)	AD (*n* = 125)
Female, *n* (%)	672 (60.81%)	197 (56.61%)	240 (66.12%)	154 (57.25%)	81 (64.80%)
Age (years)	66.60 ± 7.65	64.91 ± 6.52	66.18 ± 6.29	67.51 ± 8.70	70.55 ± 9.84
Education (years)	11.30 ± 4.24	11.92 ± 3.75	12.29 ± 3.53	9.97 ± 4.65	9.58 ± 5.17
Risk score	0.40 ± 0.33	0.21 ± 0.24	0.36 ± 0.30	0.55 ± 0.32	0.72 ± 0.23
Linear risk score	-0.76 ± 2.52	-2.18 ± 2.02	-1.03 ± 2.21	0.30 ± 2.38	1.73 ± 2.02
MMSE	26.14 ± 4.75 (*n* = 1053)	28.63 ± 1.54 (*n* = 330)	28.32 ± 1.67 (*n* = 337)	24.42 ± 3.80 (*n* = 266)	17.00 ± 5.57 (*n* = 120)
MES	24.56 ± 4.21 (*n* = 539)	26.12 ± 2.51 (*n* = 198)	25.61 ± 2.53 (*n* = 253)	20.06 ± 3.43 (*n* = 67)	11.67 ± 3.73 (*n* = 21)
MOCA-B	6.60 ± 2.82 (*n* = 540)	7.58 ± 2.12 (*n* = 198)	7.29 ± 2.12 (*n* = 254)	3.00 ± 2.08 (*n* = 69)	0.26 ± 0.81 (*n* = 19)
AVLT-L	21.64 ± 2.72 (*n* = 538)	22.49 ± 1.68 (*n* = 199)	22.31 ± 1.70 (*n* = 253)	18.51 ± 2.95 (*n* = 69)	14.41 ± 4.05 (*n* = 17)
AVLT-R	18.13 ± 4.91 (*n* = 509)	19.08 ± 4.19 (*n* = 191)	18.84 ± 4.62 (*n* = 245)	14.91 ± 3.97 (*n* = 55)	8.17 ± 3.52 (*n* = 18)
AFT	24.47 ± 3.72 (*n* = 509)	25.59 ± 2.96 (*n* = 191)	24.86 ± 3.07 (*n* = 245)	21.40 ± 3.67 (*n* = 55)	16.83 ± 5.28 (*n* = 18)
BNT	65.43 ± 28.37 (*n* = 506)	58.72 ± 18.15 (*n* = 191)	60.77 ± 21.81 (*n* = 245)	91.99 ± 34.28 (*n* = 54)	127.44 ± 58.91 (*n* = 16)
STT-A	150.41 ± 55.51 (*n* = 501)	137.95 ± 38.66 (*n* = 191)	139.47 ± 43.33 (*n* = 244)	220.77 ± 75.18 (*n* = 53)	251.85 ± 65.56 (*n* = 13)
STT-B	4.83 ± 1.48 (*n* = 377)	4.81 ± 1.59 (*n* = 132)	4.81 ± 1.41 (*n* = 199)	5.04 ± 1.39 (*n* = 34)	4.64 ± 1.70 (*n* = 12)
Plasma Aβ42/40 ratio	83.94 ± 23.90 (*n* = 382)	82.63 ± 21.90 (*n* = 133)	81.25 ± 23.05 (*n* = 203)	98.78 ± 27.49 (*n* = 34)	101.79 ± 28.68 (*n* = 12)

|					
Plasma p-tau 181 (pg/ml)	0.06 ± 0.01 (*n* = 377)	0.06 ± 0.01 (*n* = 132)	0.06 ± 0.01 (*n* = 199)	0.05 ± 0.01 (*n* = 34)	0.04 ± 0.01 (*n* = 12)
Plasma p-tau 217 (pg/ml)	2.13 ± 1.03 (*n* = 361)	1.90 ± 0.81 (*n* = 124)	1.99 ± 0.83 (*n* = 194)	3.20 ± 1.50 (*n* = 32)	3.96 ± 1.27 (*n* = 11)
Plasma *NfL* (pg/ml)	0.34 ± 0.24 (*n* = 341)	0.27 ± 0.11 (*n* = 112)	0.32 ± 0.19 (*n* = 184)	0.52 ± 0.37 (*n* = 37)	1.06 ± 0.26 (*n* = 8)
Plasma GFAP (pg/ml)	19.57 ± 13.21 (*n* = 376)	17.90 ± 12.02 (*n* = 131)	18.95 ± 12.91 (*n* = 202)	26.85 ± 17.28 (*n* = 31)	29.58 ± 9.36 (*n* = 12)

289   *Note*. All measurements are reported as mean ± s.d., unless otherwise specified. If a

290   measurement does not include all samples, the actual sample size is listed in parentheses. NC,

291   normal cognition; SCD, subjective cognitive decline; MCI, mild cognitive impairment; AD,

292   Alzheimer's disease; Risk score, MRI-based classifier's output score; Linear risk score, MRI-

293   based classifier's linear activation output score; MMSE, Mini-Mental State Examination; MES,

294   Memory Encoding Score; MOCA-B, Montreal Cognitive Assessment-Basic; AVLT-L, Auditory

295   Verbal Learning Test—Long delayed memory; AVLT-R, Auditory Verbal Learning Test—

296   Recognition; AFT, Animal Fluency Test; BNT, Boston Naming Test; STT-A, Shape Trail Test -

297   Part A; STT-B, Shape Trail Test - Part B; NfL, neurofilament light chain; GFAP, glial fibrillary

298   acidic protein.

299

AD classification performance in a Chinese sample

301   When directly applied to the SILCODE cohort without any re-training or fine-tuning, the

302   MRI-based AD classifier demonstrated robust performance in distinguishing AD from NC. As

303   **Fig. 1a** shows, the area under the curve (AUC) was 91.3% (95% CI = [88.3%, 94.0%]),

304   indicating solid discriminative capability. An optimal cutoff score of 0.38 was identified using

305   the Youden index. At this threshold, the classifier achieved an accuracy of 88.2%, a sensitivity of

306   95.2%, and a specificity of 85.6%.

307

308   **Fig. 1. Classification performance of the MRI-based AD classifier on the SILCODE**

309   **Cohort.**

310
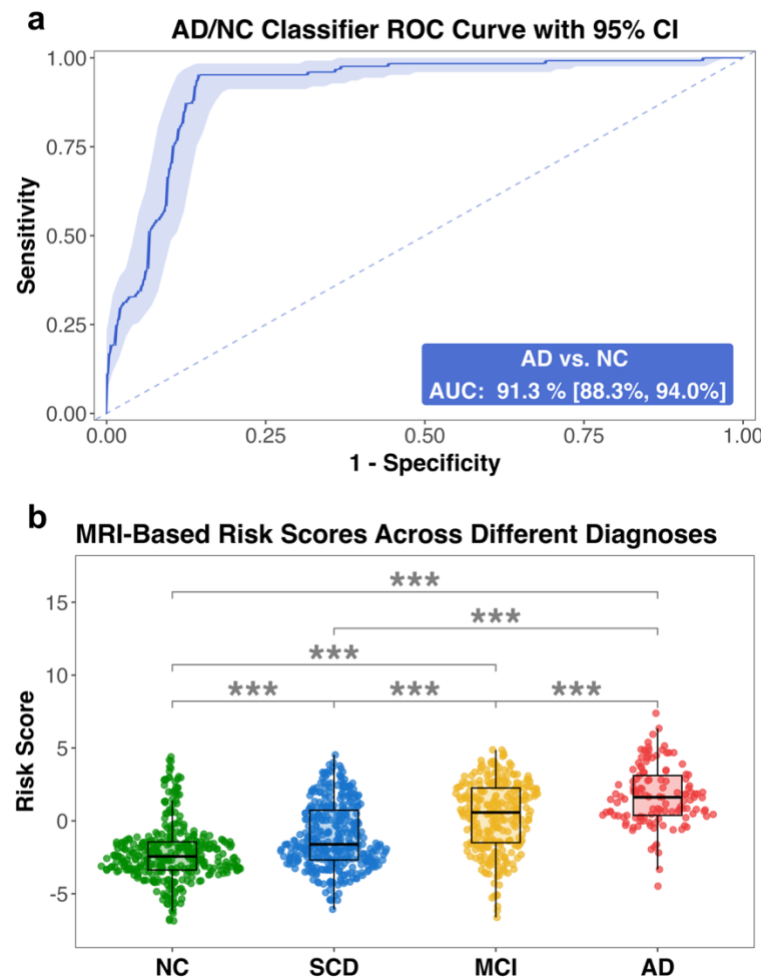
311   *Note*. **a,b,** Performance of of the MRI-based AD classifier. (a) The receiver operating

312   characteristic (ROC) curve shows the classifier's performance in distinguishing Alzheimer's

313   disease (AD) from normal cognition (NC), with an area under the curve (AUC) of 91.3% (95%

314   confidence interval [CI]: 88.3%, 94.0%). The shaded area represents the 95% CI. (b) Box plots

315   show the median, interquartile range, and whiskers extending 1.5 times the interquartile range,

316   with individual data points displayed. ***$p < 0.001$.

317

318   **Classifier scores correspond to clinical diagnoses**

319   Adjusting for sex, age, and education, and accounting for non-independent samples by

320   including participant as a random effect, we found that linear risk scores[1] varied significantly by

321   clinical diagnoses (F(3, 1058.04) = 73.96, $p < 2.20 \times 10^{-16}$). Post-hoc Tukey-corrected

322   comparisons revealed that the AD group had significantly higher scores than NC (mean

323   difference = 3.36, SE = 0.24, t(1018.86) = 14.11, $p = 7.85 \times 10^{-13}$), SCD (mean difference = 2.30,

324   SE = 0.23, t(1045.43) = 9.84, $p < 2.20 \times 10^{-16}$), and MCI groups (mean difference = 1.23, SE =

325   0.22, t(1051.45) = 5.61, $p = 1.57 \times 10^{-7}$). MCI also scored higher than NC (mean difference =

326   2.13, SE = 0.19, t(947.90) = 11.10, $p = 3.35 \times 10^{-13}$) and SCD (mean difference = 1.07, SE =

327   0.19, t(1036.81) = 5.78, $p = 6.05 \times 10^{-8}$). In addition, SCD scored higher than NC (mean

328   difference = 1.06, SE = 0.17, t(1038.16) = 6.78, $p = 6.05 \times 10^{-8}$). As shown in **Fig. 1b**, the relative

329   ordering of scores across diagnostic groups aligned with expected severity along the AD

330   continuum, despite the classifier not being trained or fine-tuned on SCD or MCI data.

331   **Performance in predicting disease progression**

332   A total of 474 participants, comprising 124 with MCI, 195 with SCD, and 155 with NC as

333   diagnosed at their baseline visits, who completed at least two visit sessions, were included in the

334   survival analysis. Participants were further classified as high-risk (risk score ≥ 0.38) or low-risk

335   (risk score < 0.38) based on their baseline sigmoid risk scores. In the full sample, 60 participants

---

[1] Model scores derived from linear activation, which are better suited for statistical testing than the default sigmoid-activated risk scores. See Methods for further details.

336   converted to AD during the 167-month follow-up period, including 57 conversions from the

337   baseline MCI group and 3 from the baseline SCD group. No baseline NC participant converted

338   to AD during the tracking period. Kaplan-Meier survival plots illustrate the detailed progression

339   of AD conversion in the full sample (**Fig. 2a**), baseline MCI group (**Fig. 2b**), and baseline SCD

340   group (**Fig. 2c**). Notably, as shown in the table at the bottom of **Fig. 2a**, our model identified 52

341   of 60 AD converters as high risk at baseline, achieving an overall capture rate of 86.7%. In the

342   full sample, 52 of the 177 participants classified as high-risk and 8 of the 297 participants

343   classified as low-risk converted to AD during the follow-up period. The log-rank test indicated a

344   significant difference between the survival curves ($p < 0.0001$), accounted for by higher

345   probability of conversion for participants in the high-risk group.

346   Specifically, among participants with MCI at baseline (**Fig. 2b**), 49 of the 82 high-risk

347   participants converted to AD, compared to 8 of the 42 low-risk participants, showing a

348   significantly higher probability of conversion to AD in the high-risk group (log-rank test, $p <$

349   0.0001). The median survival time (i.e., time-to-conversion) of the high-risk MCI group (31

350   months) was significantly shorter than that of the low-risk MCI group (>167 months). In

351   participants with SCD at baseline, only 3 conversions occurred, and all were in the high-risk

352   category (**Fig. 3c**). The log-rank test indicated a significantly higher risk of conversion for high-

353   risk SCD participants ($p = 0.027$, **Fig. 2**). In summary, survival analyses demonstrated that

354   participants identified as high-risk at baseline by our classifier were more likely to convert to AD

355   over time.

356

357

358

359 **Fig. 2. Survival analysis for baseline MCI and SCD groups.**


360
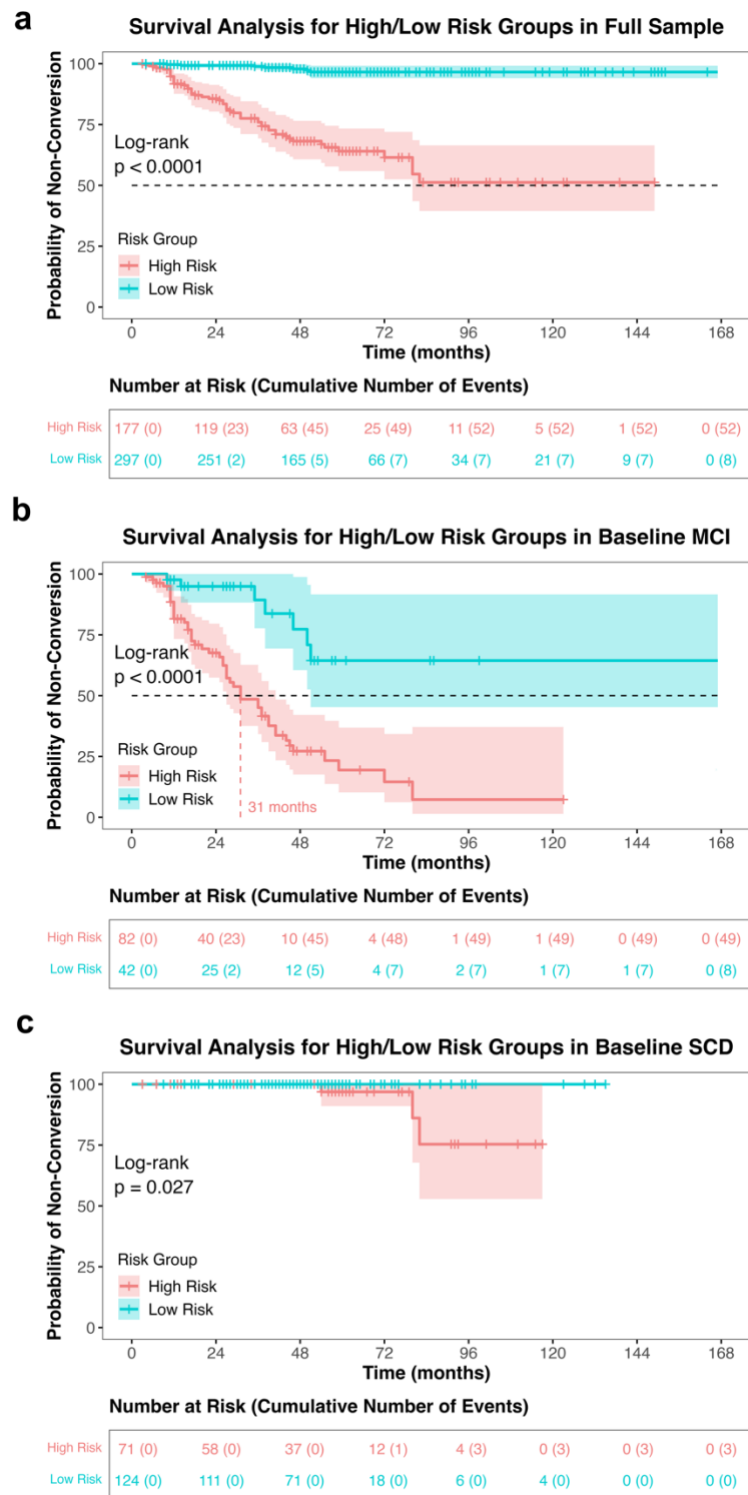
361 *Note.* **a, b, c,** Kaplan-Meier survival (conversion) curves showing the probability of non-

362 conversion in high-risk (red) and low-risk (blue) groups over 167 months (the longest follow-up

363 period). High risk was defined by baseline AD sigmoid risk scores (≥ 0.38). Curves are shown for

364 the full sample (**a**), MCI (**b**), and SCD (**c**) at baseline. The baseline NC sample is not shown, as

365 no baseline NC participant progressed to AD during the tracking period. Shaded areas represent

366 95% confidence intervals, with vertical ticks indicating censored participants (lost to follow-up).

367 Vertical dotted lines mark median conversion times, and the horizontal black dashed line

368 indicates a 50% probability of non-conversion. Log-rank tests compared survival (conversion)

369 distributions between the two groups, with $p < 0.05$ indicating significance. Tables below the

370 plots show the number of participants at risk (i.e., still followed and non-progressed) at each time

371 point, with cumulative conversion events in parentheses. To avoid confusion, the survival curves

372 for the low-risk group (**a** and **b**) omit the abrupt drop to zero at the tail caused by the last single

373 individual's conversion to AD. MCI, mild cognitive impairment; SCD, subjective cognitive

374 decline.

375

Classifier predicted AD progression years before diagnosis

377   Using all visit sessions ($n = 88$) from the 60 participants who progressed from NC, SCD, or

378 MCI, we estimated *how early* the classifier was able to predict AD conversion. Specifically, we

379 calculated the ratio of sessions flagged as high risk by the classifier (risk score ≥ 0.38) at various

380 time intervals before the first AD diagnosis. Results (**Fig. 3a**) show that the classifier achieved a

381 capture rate of about 90% within 3 years, 78% within 3–5 years, and 57% within 5–14 years

382 before the first AD diagnosis among AD converters. Mean risk scores increased monotonically

383 across follow-up periods, consistent with the probability of conversion (**Fig. 3b**).

Fig. 3. Proportion exceeding high-risk criteria and mean sigmoid risk scores before AD diagnosis.
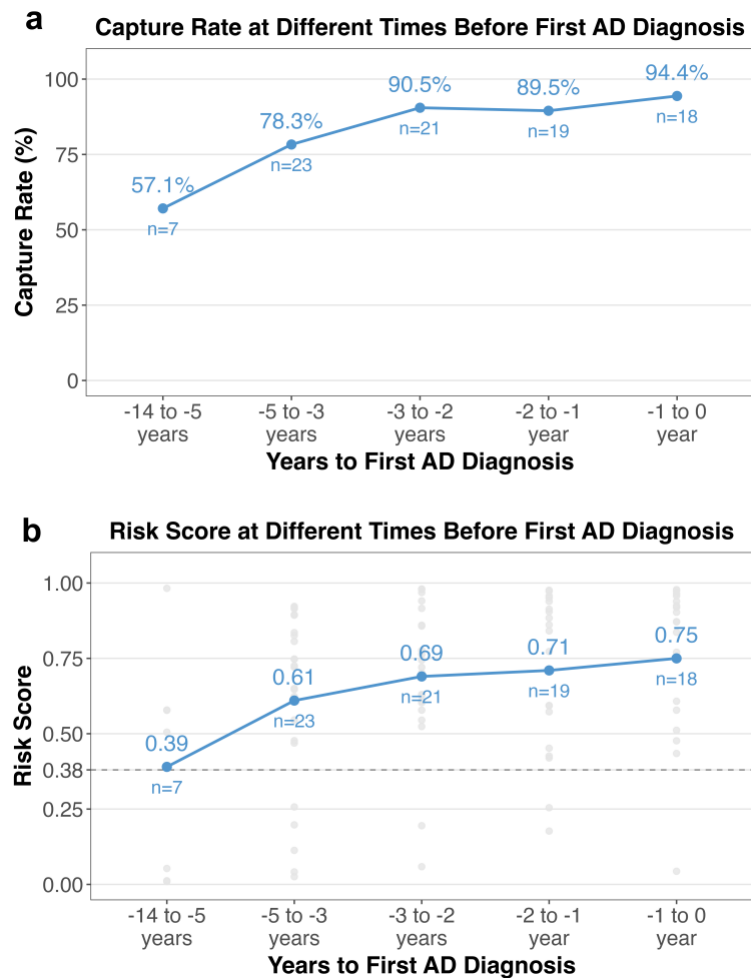
Note. **a**, **b**, The proportion of sessions with sigmoid risk scores meeting the high-risk threshold (≥ 0.38) (**a**) and the mean sigmoid risk scores (**b**) across different time intervals before the first AD diagnosis. The numbers in **a** represent the proportions, while the numbers in **b** indicate the mean risk scores. *n* refers to the number of available sessions in each time period. The horizontal dotted grey line in **b** marks the 0.38 cut-off for high risk. AD, Alzheimer's disease.

Risk scores associated with cognitive assessments and plasma biomarkers

Linear mixed-effects models were employed to examine the associations of MRI-based AD

risk scores with standardized cognitive assessments and plasma biomarker levels. Global

associations were first assessed across all samples, without distinguishing diagnostic phases, to

identify general patterns. Diagnosis-specific analyses were then conducted within each pre-AD

diagnostic phase to evaluate more granular associations, given the importance of these phases for

understanding early-stage changes critical for detection and intervention.

As shown in **Fig. 4a** and **Table 2**, MRI-based risk scores were significantly associated with

all cognitive measures, including both global cognitive assessments (i.e., MMSE, MES, MOCA-

B) and subdomain-specific tests (i.e., AVLT-L, AVLT-R, AFT, BNT, STT-A, STT-B). Higher

MRI-based risk scores were linked to poorer cognitive performance across the board.

Specifically, higher scores on the STT-A, and STT-B (where higher scores indicate worse

cognitive function) were associated with higher risk scores ($ps < 0.01$, $f^2s = 0.02$). By contrast,

for the remaining cognitive measures, where lower scores indicate worse cognitive function (e.g.,

MMSE, MES, MOCA-B), higher risk scores were associated with lower scores, indicating a

decline in cognitive performance ($ps < 0.01$, $f^2s = 0.02 - 0.10$). When examined within each

diagnostic phase, significant associations between MRI-based risk scores and cognitive

performance were observed only within the MCI phase. Specifically, risk scores remained

significantly associated with MMSE ($p < 0.001$, $f^2 = 0.04$) and MOCA-B ($p < 0.05$, $f^2 = 0.05$)

within the MCI phase.

When plasma biomarkers were examined across all samples (**Fig. 4c** and **Table 2**), higher

MRI-based risk scores were positively associated with elevated levels of p-tau 181, p-tau 217,

NfL, and GFAP ($ps < 0.05$, $f^2s = 0.01 - 0.02$), indicating that higher risk scores are linked to

415 increased levels of these neurodegenerative markers. When examined by diagnosis, significant

416 associations between MRI-based risk scores and plasma biomarker levels were observed only

417 within the NC phase. Specifically, risk scores remained significantly associated with p-tau 217 (p

418 $< 0.05, f^2 = 0.03$) and NfL ($p < 0.05, f^2 = 0.04$). In the SCD and MCI groups, the associations

419 were all non-significant (SCD: $ps > 0.05, f^2 s \approx 0$; MCI: $ps > 0.05, f^2 s = 0$–0.02). The associations

420 between risk scores and the plasma Aβ42/40 ratio were non-significant, both across all samples

421 and within diagnoses.

422

423 **Fig. 4. Risk score associations with cognitive performance and plasma biomarkers.**

424
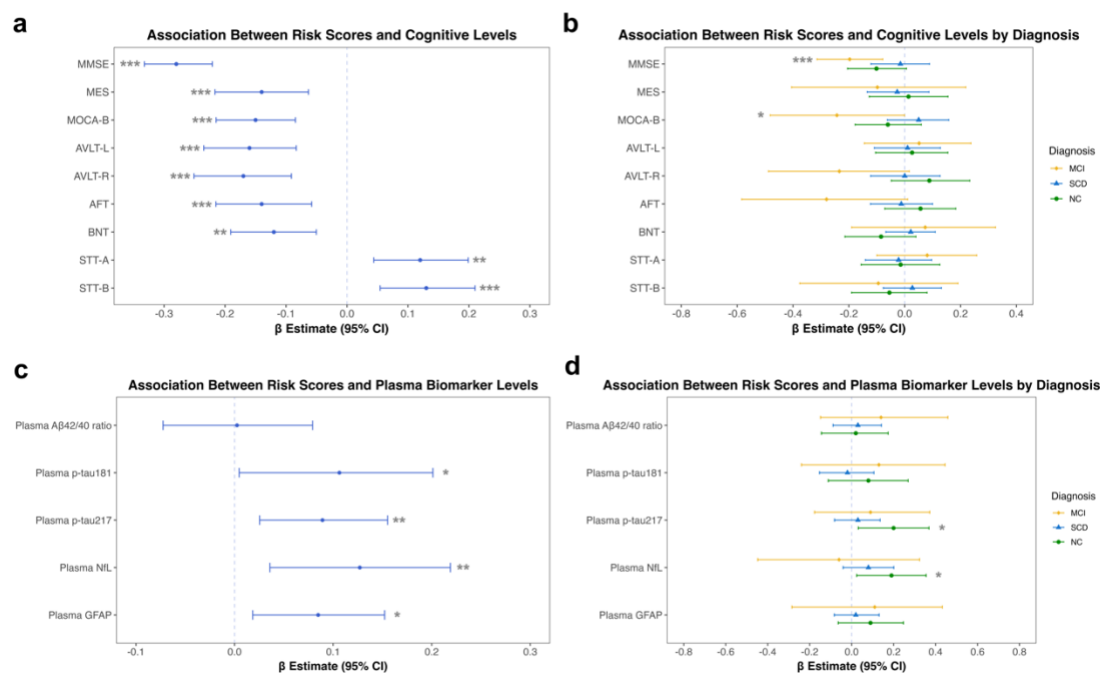
425 **a, b,** Associations between MRI-based risk scores and cognitive performance across all samples

426 (**a**) and within each pre-AD group (**b**). **c, d,** Associations between risk scores and plasma

427 biomarker levels across all samples (**c**) and within each pre-AD group (**d**). Separate models were

428 fitted for each cognitive and plasma measure, with all variables standardized. The linear risk

429 score was the independent variable, while sex, age, and education were included as covariates.

430   Most models used linear mixed-effects analysis, with participant included as a random effect.

431   However, for plasma measures within the MCI group, general linear models without random

432   effect were used instead due to the limited number of repeated measures per participant, which

433   caused convergence issues in mixed-effects models. Horizontal lines represent 95% confidence

434   intervals (CIs; based on 5,000 bootstrap samples) for the β estimates of the linear risk score; CIs

435   that do not include 0 indicate statistical significance. The dotted vertical line represents zero,

436   with estimates to the left indicating negative associations and those to the right indicating

437   positive associations. NC, normal cognition; SCD, subjective cognitive decline; MCI, mild

438   cognitive impairment; AD, Alzheimer's disease; CI, confidence interval; MMSE, Mini-Mental

439   State Examination; MES, Memory Encoding Score; MOCA-B, Montreal Cognitive Assessment-

440   Basic; AVLT-L, Auditory Verbal Learning Test—Long delayed memory; AVLT-R, Auditory

441   Verbal Learning Test— Recognition; AFT, Animal Fluency Test; BNT, Boston Naming Test;

442   STT-A, Shape Trail Test - Part A; STT-B, Shape Trail Test - Part B; NfL, neurofilament light

443   chain; GFAP, glial fibrillary acidic protein. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 2. Associations between MRI-Based Risk Scores and Cognitive Measures and Plasma Biomarkers, across All Samples and by Diagnostic Phase.

Measure		All Samples								NC						
Measure	n	β	f^2	SE	df	t	p	95% CI	n	β	f^2	SE	df	t	p	95% CI
MMSE	1053	-0.28***	0.10	0.03	1032.52	-9.97	2.03×10^{-22}	[-0.33, -0.22]	330	-0.1	0.01	0.05	320.2	-1.89	0.059	[-0.2, 0.01]
MES	492	-0.14***	0.02	0.04	386.51	-3.63	3.27×10^{-4}	[-0.22, -0.06]	188	0.01	0	0.07	181.54	0.19	0.849	[-0.13, 0.15]
MOCA-B	539	-0.15***	0.03	0.03	373.46	-4.55	7.24×10^{-6}	[-0.21, -0.08]	198	-0.06	0	0.06	149.24	-1.00	0.321	[-0.18, 0.06]
AVLT-L	540	-0.16***	0.03	0.04	512.52	-4.07	5.38×10^{-5}	[-0.23, -0.08]	198	0.03	0	0.07	177.44	0.40	0.689	[-0.1, 0.15]
AVLT-R	538	-0.17***	0.03	0.04	512.44	-4.26	2.49×10^{-5}	[-0.25, -0.09]	199	0.09	0.01	0.07	180.21	1.23	0.221	[-0.05, 0.23]
AFT	509	-0.14***	0.02	0.04	471.65	-3.37	8.14×10^{-4}	[-0.22, -0.06]	191	0.06	0	0.06	142.31	0.88	0.379	[-0.07, 0.18]
BNT	509	-0.12**	0.02	0.04	378.24	-3.3	0.001	[-0.19, -0.05]	191	-0.08	0.01	0.07	155.98	-1.30	0.195	[-0.21, 0.04]
STT-A	506	0.12**	0.02	0.04	459.61	3.01	0.003	[0.04, 0.20]	191	-0.01	0	0.07	179.16	-0.20	0.840	[-0.16, 0.13]
STT-B	501	0.13***	0.02	0.04	451.06	3.4	7.22×10^{-4}	[0.05, 0.21]	191	-0.05	0	0.07	179.31	-0.80	0.423	[-0.19, 0.08]
Plasma Aβ42/40 ratio	377	0.003	0	0.04	215.68	0.06	0.949	[-0.07, 0.08]	132	0.02	0	0.08	74.89	0.22	0.824	[-0.14, 0.17]
Plasma p-tau 181	361	0.11*	0.01	0.05	328.24	2.22	0.027	[0.00, 0.20]	124	0.08	0.01	0.1	118.35	0.83	0.410	[-0.11, 0.27]
Plasma p-tau 217	341	0.09**	0.01	0.03	164.86	2.72	0.007	[0.03, 0.16]	112	0.2*	0.03	0.08	76	2.39	0.019	[0.03, 0.37]

	n	β	f²	SE	df	t	p	95% CI	n	β	f²	SE	df	t	p	95% CI
Plasma NfL	**376**	0.13**	0.02	0.05	359.88	2.73	0.007	[0.04, 0.22]	**131**	0.19*	0.04	0.08	126	2.22	0.028	[0.02, 0.35]
Plasma GFAP	**379**	0.08*	0.01	0.03	198.03	2.47	0.014	[0.02, 0.15]	**132**	0.09	0	0.08	90.32	1.16	0.248	[-0.06, 0.25]

	SCD								*MCI*							
Measure	*n*	*β*	*f²*	*SE*	*df*	*t*	*p*	*95% CI*	*n*	*β*	*f²*	*SE*	*df*	*t*	*p*	*95% CI*
MMSE	**337**	-0.02	0	0.05	311.05	-0.28	0.778	[-0.12, 0.09]	**266**	-0.2***	0.04	0.06	259.94	-3.35	9.28×10^{-4}	[-0.31, -0.08]
MES	**241**	-0.03	0	0.06	230.02	-0.47	0.642	[-0.13, 0.09]	**49**	-0.1	0.01	0.15	29.94	-0.66	0.517	[-0.41, 0.22]
MOCA-B	**253**	0.05	0	0.05	246.27	0.91	0.363	[-0.06, 0.16]	**67**	-0.24*	0.05	0.12	50.47	-2.05	0.046	[-0.48, 0]
AVLT-L	**254**	0.01	0	0.06	248.8	0.17	0.866	[-0.11, 0.13]	**69**	0.05	0	0.09	24.39	0.56	0.582	[-0.14, 0.24]
AVLT-R	**253**	0	0	0.06	247.06	0	0.999	[-0.12, 0.13]	**69**	-0.23	0.04	0.13	57.28	-1.87	0.067	[-0.49, 0.02]
AFT	**245**	-0.01	0	0.06	231.12	-0.22	0.828	[-0.12, 0.1]	**55**	-0.28	0.06	0.15	42.66	-1.92	0.061	[-0.58, 0.01]
BNT	**245**	0.02	0	0.05	150.82	0.47	0.638	[-0.07, 0.11]	**55**	0.07	0	0.13	47.97	0.57	0.574	[-0.19, 0.33]
STT-A	**245**	-0.02	0	0.06	238.83	-0.36	0.716	[-0.14, 0.1]	**54**	0.08	0	0.09	11.49	0.91	0.382	[-0.1, 0.26]
STT-B	**244**	0.03	0	0.05	211.8	0.53	0.598	[-0.08, 0.13]	**53**	-0.09	0.01	0.14	44.03	-0.67	0.503	[-0.38, 0.19]
Plasma Aβ42/40 ratio	**199**	0.03	0	0.06	130.14	0.46	0.649	[-0.09, 0.14]	**34**	0.14	0.02	0.17	29	0.85	0.404	[-0.15, 0.46]
Plasma p-tau 181	**194**	-0.02	0	0.07	188.23	-0.35	0.725	[-0.15, 0.11]	**32**	0.13	0.02	0.19	27	0.72	0.476	[-0.24, 0.45]
Plasma p-tau 217	**184**	0.03	0	0.05	118.18	0.48	0.632	[-0.08, 0.14]	**37**	0.09	0.01	0.14	32	0.64	0.528	[-0.18, 0.37]

Plasma NfL	202	0.08	0.01	0.06	184.01	1.3	0.195	[-0.04, 0.2]	31	-0.06	0	0.2	26	-0.28	0.783	[-0.45, 0.32]
Plasma GFAP	203	0.02	0	0.05	140.09	0.44	0.659	[-0.08, 0.13]	32	0.11	0.02	0.16	27	0.66	0.513	[-0.28, 0.43]

Note. Associations of MRI-based risk scores with cognitive measures and plasma biomarkers were examined across all samples and within each diagnostic group. β represents the regression coefficient of the linear risk score from each model, with all variables standardized for easier comparison. f^2 is Cohen's f^2, a measure of effect size that indicates the extent to which the inclusion of a specific predictor variable improves the explanatory power of the model. All models adjusted for sex, age, and education. Most models included participants as a random effect, except for plasma measures within the MCI group, where general linear models without random effect were used to avoid convergence issues. 95% CI was computed using 5,000 bootstrap iterations. NC, normal cognition; SCD, subjective cognitive decline; MCI, mild cognitive impairment; SE, standard error of the estimate; df, degrees of freedom; CI, confidence interval. MMSE, Mini-Mental State Examination; MES, Memory Encoding Score; MOCA-B, Montreal Cognitive Assessment-Basic; AVLT-L, Auditory Verbal Learning Test—Long delayed memory; AVLT-R, Auditory Verbal Learning Test— Recognition; AFT, Animal Fluency Test; BNT, Boston Naming Test; STT-A, Shape Trail Test - Part A; STT-B, Shape Trail Test - Part B; NfL, neurofilament light chain; GFAP, glial fibrillary acidic protein. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

457 **Subtyping and predicting cognitive decline speed via an interpretable model**

458   Individualized AD brain risk maps were generated for each MRI session for all participants.

459 In the AD group, all patients could be clustered into three categories based on the distribution of

460 the risk areas, forming three MRI-based AD subtypes (**Fig. S2**). The essential high-risk brain

461 regions of the three AD subtypes are shown in **Fig. 5a-c**. The primary risk regions for AD

462 subtype 1 were located around the midbrain area, including the midbrain tectum, red nucleus,

463 dorsal raphe nucleus, and pineal gland. The primary risk regions for AD subtype 2 included

464 hippocampus and posterior cingulate cortex. The primary risk regions for AD subtype 3 included

465 hippocampus, the midbrain tectum, red nucleus, dorsal raphe nucleus, and pineal gland. All

466 SILCODE AD patients misclassified as NC by the MRI-based classifier belong to AD subtype 1

467 (**Fig. 5d**).

468

469

470

471

472

473

474

475

476

477

**Fig. 5. Brain risk map-based Alzheimer's disease subtypes.**
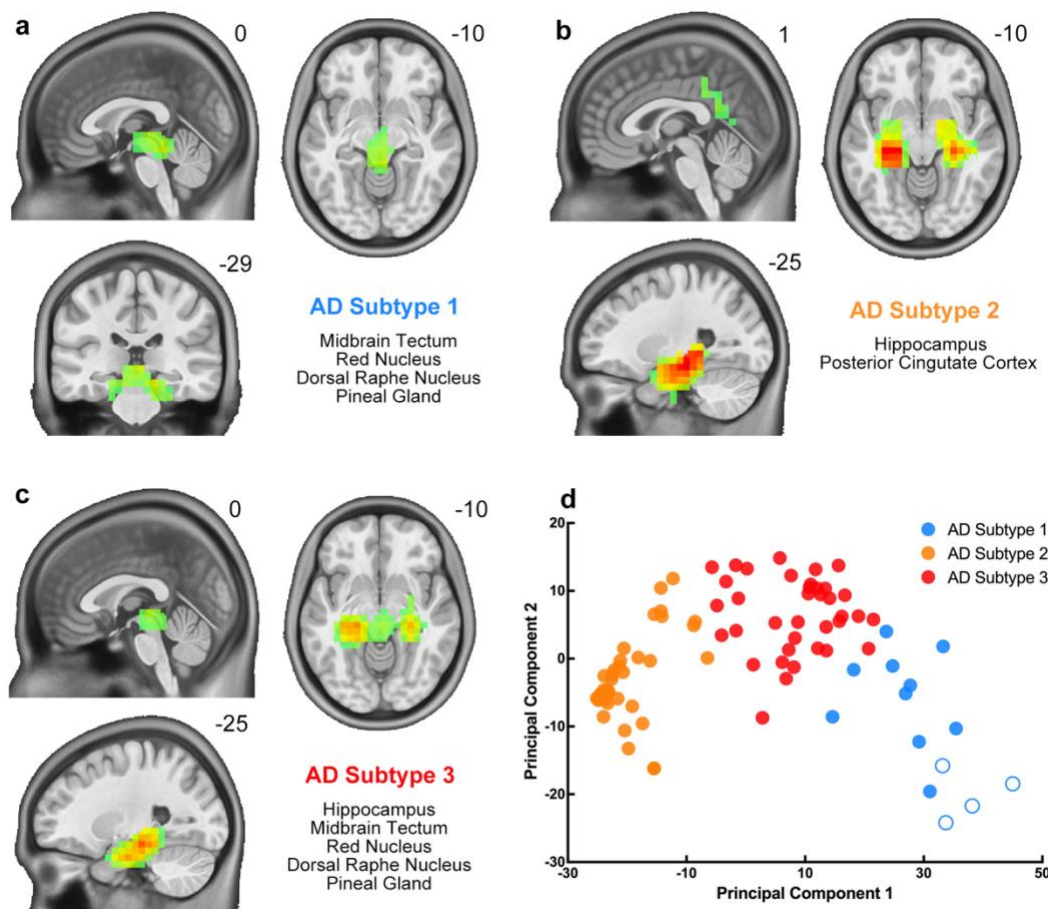
Note. **a-c,** Distribution of brain regions with the highest disease risk for the three AD subtypes.

The brain anatomical structures with the highest risks are listed in the bottom right of each of the

panels. **d,** Patients of the three AD subtypes in a two-dimensional distribution map based on

principal component analysis. Solid circles represent patients with high-risk MRI-based AD

classifier scores, while hollow circles represent patients with low-risk MRI-based AD classifier

scores. AD, Alzheimer's disease.

To assess the potential of brain subtypes to predict disease progression, we compared the

rate of cognitive decline, measured as the annual change in MMSE score (the difference between

489   baseline and the most recent score divided by the number of years), across brain subtypes. We

490   focused on individuals with baseline MCI diagnoses ($n = 111$), as this phase typically shows the

491   most pronounced cognitive deterioration. Results indicate that cognitive decline rates differed

492   significantly among brain subtypes in the MCI group, even after adjusting for sex, age, and

493   education ($F(2, 105) = 3.63$, $p = 0.030$; **Fig. 6b**). Post-hoc comparisons revealed that MCI

494   patients classified as subtype 3 experienced a faster cognitive decline compared to those in

495   subtype 1 (mean difference = 1.46, SE = 0.69, $t(105) = 2.12$, $p = 0.036$) and 2 (mean difference =

496   1.52, SE = 0.70, $t(105) = 2.17$, $p = 0.032$). In contrast, no significant differences were observed

497   in baseline MMSE scores across the brain subtypes ($F(2,105) = 1.56$, $p = 0.212$; **Fig. 6a**).

498   Demographics and risk score profiles for different brain subtypes in the MCI group are presented

499   in **Fig. S3**. Cognitive decline rate did not differ significantly among subtypes in the NC or SCD

500   groups (**Fig. S4-5**)

501

502

503

504

505

506

507

508

509

510   **Fig. 6. Baseline scores and change rates of MMSE scores across different brain subtype**

511   **groups in individuals with baseline MCI.**

512
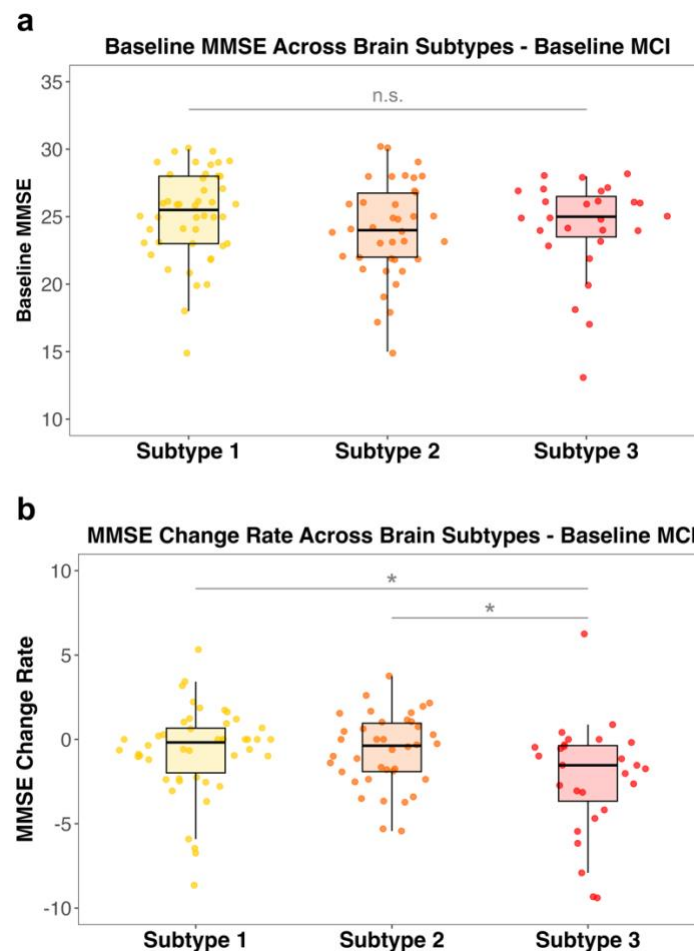
513   *Note*. **a,** Baseline MMSE scores across different brain subtype groups in individuals with

514   baseline MCI; **b,** MMSE change rates across different brain subtype groups in individuals with

515   baseline MCI. MMSE, Mini-Mental State Examination; MCI, mild cognitive impairment. n.s.,

516   non-significant, *$p < 0.05$.

517

DISCUSSION

We examined the potential clinical utility of an MRI-based AD deep learning model within a comprehensively phenotyped longitudinal Chinese cohort. The classifier, which had been trained on North American samples, demonstrated good generalizability in distinguishing AD from NC within the Chinese sample, achieving an AUC of 91.3% with high sensitivity. Importantly, the model accurately identified 86.7% of participants who would progress to AD up to 14 years before their AD diagnosis. MCI patients defined as high risk by the model were significantly more likely to progress to AD. The interpretable model provided personalized brain risk maps and identified the MCI subtype that is most prone to rapid cognitive decline. These predictive abilities enable the use of structural MRI to select patients who are most likely to benefit from early drug intervention. Finally, the MRI-based risk scores were significantly correlated with global cognitive measurements and plasma-based biomarkers.

AD classification performance and generalizability

The classifier's ability to accurately distinguish AD from NC in the SILCODE Chinese cohort and previously in four Western samples underscores its robustness and potential for broader application across diverse populations[34]. Currently, MRI is primarily used for staging, monitoring disease progression, and detecting potential adverse events from pharmacological treatments, such as amyloid related imaging abnormalities (ARIA). However, MRI integrated with artificial intelligence is poised to be more closely integrated into screening, early diagnostics and precision interventions for the AD spectrum[45]. Compared to PET scans and cerebrospinal fluid tests, MRI's non-invasive and radiation-free nature, ready availability, low cost and high sensitivity make it more suitable for large-scale disease screening[46]. This may facilitate the decision-making process for proceeding with monoclonal antibody treatments,

541   thereby significantly advancing the timing and potentially enhancing the effectiveness of

542   interventions. The classifier's slightly lower performance (AUC of 91.3%) than in the original

543   study is likely due to differences in population characteristics. Age is the most critical factor

544   influencing the model's predictive performance. The samples used for training and testing the

545   MRI-based models in the previous study (76 ± 7 years in ADNI, 74 ± 7 years in AIBL, 70 ± 7

546   years in MIRIAD and 75 ± 10 years in OASIS)[17] were significantly older than the SILCODE

547   samples (66.6 ± 7.6 years). According to the 2024 report by the Alzheimer's Association, the

548   incidence of AD in the 75-84 age group is 2.64 times higher than in the 65-74 age group[47]. This

549   implies that a considerable number of SILCODE participants with high-risk scores, currently

550   diagnosed as NC, are likely to progress to AD in the future. In fact, this inference aligns with the

551   model's performance, which demonstrates a remarkable sensitivity of 95.2% but a less robust

552   specificity of 85.6%. These findings underscore the model's effectiveness in correctly detecting

553   AD, making it well-suited for large scale AD screening and selecting patients for drug therapies.

554   In such scenarios, the priority is to minimize the likelihood of overlooking AD patients and those

555   at high risk for developing AD. Furthermore, considering the classifier can provide an in-

556   advance alert of AD progression risk, it is plausible that at least some of the healthy subjects

557   currently identified as "false positives" might be at a higher risk of developing AD in the future,

558   warranting closer monitoring.

559   In addition, structural differences in brain gray matter volume and density between

560   Caucasian and East Asian populations[48,49] may affect classifier performance. Given the limited

561   number of AD patients in SILCODE, we chose not to fine-tune the model so as to avoid

562   overfitting. Nevertheless, testing the model's generalizability directly in a different population

563   could also be a valuable next step. Furthermore, although the current model was not trained with

564 SCD or MCI patients, the output scores for different clinical diagnoses differed significantly.

565 Higher risk scores corresponded to more advanced clinical stages, highlighting a valuable

566 gradient of AD-related effects.

567 **Predicting early-stage AD progression by MRI-based risk scores**

568 The results of the survival analysis and lead time analysis further validated that MRI

569 combined with deep learning could effectively predict the progression of early-stage AD

570 patients. Previous research has shown that structural MRI changes can be detected ten years

571 before clinical diagnosis at the group level [50,51]. In the current study, the classifier was able to

572 detect abnormal scores up to 14 years before diagnosis at the individual level—with 57% of

573 converters showing abnormal scores 5–14 years prior, 78% within 3–5 years, and ~90% within 3

574 years. This indicates that the current model can detect early-stage neurodegenerative changes

575 well before the onset of significant cognitive decline and visually observable brain atrophy.

576 Furthermore, the MCI patients identified as high-risk by the MRI-based risk score showed a

577 significantly higher probability of progressing to AD compared to those classified as low-risk.

578 The median survival time (i.e., time-to-conversion) for the high-risk MCI group (31 months) was

579 significantly shorter than that of the low-risk MCI group (>167 months). This provides a

580 valuable tool to identify high-risk MCI patients who are more likely to benefit from close

581 monitoring and intervention. Emerging treatments, such as those targeting amyloid plaques and

582 tau pathology, are more effective when applied during the earliest phases of

583 neurodegeneration[52,53]. By identifying high-risk individuals early, clinicians can intervene at a

584 stage when disease-modifying therapies may have the greatest impact, potentially delaying AD

585 progression and improving long-term outcomes.

586

Subtyping and predicting rate of cognitive decline via an interpretable model

The individualized brain risk maps derived from our interpretable models offer a novel

approach to understand and visualize disease pathology and progression at the individual level.

The ability to identify specific brain regions affected by AD may facilitate targeted interventions,

i.e., personalized medicine, and enhance physician confidence in the deep learning model.

Interestingly, the AD-related risk areas include not only the most commonly affected

hippocampus and posterior cingulate cortex, but also encompass many deep structures such as

the midbrain tectum, red nucleus, dorsal raphe nucleus, and pineal gland. These brain regions

constitute the main risk areas corresponding to Subtype 1. These regions are involved in the

processing of audiovisual information and motor control, and regulate circadian rhythms and

mood through the secretion of melatonin and serotonin[54–58]. A substantial amount of

morphological and pathological evidence shows that these structures are altered in AD, and are

closely associated with symptoms such as sleep disturbances, audiovisual integration, and

depressed mood. For example, the RNA binding protein, TDP-43, is strongly associated with the

morphological and clinical features of AD, and TDP-43 deposited in midbrain tectum is included

as a key feature of Stage 5 in the latest AD staging scheme[59,60]. Additionally, the pineal gland in

AD patients is significantly smaller than in NC and MCI and its volume is associated with

cognitive decline[58]. Furthermore, the subtypes based on the risk map can be used to identify MCI

patients with the fastest cognitive decline (Subtype 3). This can provide guidance for selecting

MCI patients who are most likely to benefit from pharmacological treatment.

Physiological and cognitive relevance of the MRI-based risk scores

We also examined the relationship between classifier scores and clinical measurement,

including blood biomarkers and cognitive assessments. We used plasma biomarkers from

610   SILCODE due to their greater accessibility in practice than CSF markers. We used the A-T-N-I

611   framework, which stands for Amyloid (Aβ 42/40 ratio), Tau (p-tau 181 and p-tau 217),

612   Neurodegeneration (NfL), and Inflammation (GAFP), to evaluate the physiological relevance of

613   the MRI-based risk score. Among blood biomarkers, we detected significant global correlations

614   between the risk score and the T, N, and I biomarkers. This demonstrates that the classifier

615   scores are not only indicative of structural brain changes observable via MRI but also correspond

616   to biochemical changes occurring in blood.

617   When analyzing the relationship between risk scores and plasma biomarkers within

618   diagnostic groups, significant associations were observed only in the NC group. This is not

619   surprising, as plasma biomarkers change most rapidly during the preclinical stage[61]. Notably, the

620   association between the risk score and p-tau 181became non-significant when tested within the

621   stages, while the association with p-tau 217 remained significant in the NC stage. This finding

622   aligns with recent studies that identify plasma p-tau 217 as the most accurate and robust

623   biomarker for distinguishing preclinical pathological stages[62] and predicting the progression of

624   AD among phosphorylated tau markers[63]. Importantly, in the 2024 revision of the NIA-AA

625   criteria, p-tau 217 is the sole plasma biomarker utilized for diagnostic purposes and the only

626   plasma biomarker that is used simultaneously for diagnosis, staging, monitoring progression, and

627   assessing treatment efficacy[20]. The remarkable correlation between the MRI-based risk score and

628   plasma p-tau 217 further validates the diagnostic and screening value of MRI. The significant

629   correlation of MRI-based risk scores with NfL scores is not surprising because brain anatomical

630   structures and NfL levels show concurrent alterations, both manifesting in parallel along the AD

631   progression trajectory[4,5]. The significance of NfL within NC groups aligns with evidence

632 supporting it as a risk biomarker for predicting hippocampal volume changes, a key feature of

633 AD pathology, in cognitively unimpaired individuals[64].

634 Interestingly, the amyloid biomarker did not correlate significantly with model scores. This

635 may be because plasma Aβ is not a sufficiently strong biomarker for AD. Supporting this, **Table**

636 **1** shows that plasma Aβ42/40 ratios were nearly identical across different diagnoses. Prior

637 studies have similarly reported no significant differences in plasma Aβ42/40 across AD stages

638 and found only weak correlations between plasma Aβ42/40 and its CSF or PET counterparts[65,66].

639 Differences in methods for measuring plasma Aβ can also impact results. A study comparing

640 various methods and assays found that plasma Aβ42/40 levels measured by the Simoa

641 immunoassay[67], particularly from Quanterix, as used in the current dataset, do not adequately

642 represent CSF Aβ42/40 or Aβ PET status. Future research should consider using CSF Aβ data or

643 Aβ PET to test the correlation between Aβ and risk scores from our AD classifier, or employ

644 plasma methods that better reflect CSF Aβ levels, such as the immunoprecipitation-coupled mass

645 spectrometry (IP-MS) method[67].

646 Although the significant correlations between the MRI-based risk scores and the plasma

647 biomarkers primarily occur in NC samples, the current study cannot entirely rule out the

648 possibility that these associations also occur in MCI population. The small sample size of the

649 MCI individuals with both plasma and MRI data (approximately 30) reduced the likelihood of

650 detecting significant associations.

651 Associations of MRI-based risk scores with both global cognitive measures (MMSE, CDR,

652 MES, MOCA-B) and subdomain-specific tests (AVLT-L, AVLT-R, AFT, BNT, STT-A, STT-B)

653 were significant, demonstrating the consistency between brain risk and cognitive function.

654 Within diagnosis, significant associations were observed only during the MCI phase, where

655   cognitive symptoms deteriorate significantly[61]. The remaining significant associations of risk

656   scores with MMSE and MOCA-B, two widely used global cognitive measures in clinical

657   settings, support the clinical applicability of the risk score and suggest that MRI-based risk

658   scores are more closely related to global cognitive status than to specific cognitive subdomains.

659   **Strength and limitations**

660   The main strength of this study involves the use of a large, longitudinal, well-phenotyped

661   dataset that is independent of the training samples to validate the ability of the MRI-based deep

662   learning model to predict disease progression and cognitive decline. The non-invasive, cost-

663   effective, and widely accessible nature of MRI support expanding its role in AD drug

664   interventions, from safety monitoring for adverse events such as ARIA[68] to selecting high-

665   priority patients and monitoring drug efficacy. The physiological significance of the MRI-based

666   risk scores is validated by widely accepted and accessible blood-based biomarkers. A few

667   machine learning or deep learning studies have linked MRI-based AD scores with molecular

668   pathology, such as CSF amyloid or tau pathology[69,70]. The current study contributes to this field

669   by providing an evaluation of A-T-N-I pathology in plasma. In the field of neuroregulation, the

670   individualized brain risk maps in this study could help researchers and clinicians pinpoint

671   potential intervention targets. The model has been made open-source and integrated into a free,

672   publicly accessible online platform designed to revolutionize early screening and intervention for

673   AD. By openly sharing our developed models (https://github.com/Chaogan-

674   Yan/BrainImageNet), we empower researchers worldwide to adopt and deploy these tools,

675   fostering global collaboration in AD research. Furthermore, our demonstration website

676   (http://brainimagenet.org) enables users to upload raw T1-weighted MRI scans or preprocessed

677   gray matter density (GMD) and gray matter volume (GMV) data. The platform delivers real-time

678 predictions of AD risk, providing an invaluable resource for clinicians and researchers to identify

679 at-risk individuals early and explore tailored intervention strategies. This initiative underscores

680 our commitment to leveraging advanced technology for proactive, accessible, and scalable AD

681 risk assessment.

682 Study limitations include the use of a single-site dataset, SILCODE, which hinders effective

683 simultaneous fine-tuning and testing of the model. Future studies should assemble a more

684 extensive and representative multi-center AD dataset to fine-tune the current MRI-based models

685 for better generalizability and to further enhance its value in clinical application. Moreover, the

686 average age of our sample was around 66, which is lower than comparable studies. A future

687 research goal is to build a normative model based on a large sample of healthy individuals to

688 construct age-specific models to improve specificity.

689 In summary, our MRI-based deep learning AD classifier generalized to a different

690 population, identifying 87% of participants at baseline who progressed to AD. The interpretable

691 models indicated AD risk brain regions and predicted the MCI patients who experienced the

692 fastest cognitive declines. Implementation of MRI in clinical practice can complement current

693 biomarkers such as CSF, PET, and blood tests, facilitating the selection of patients most likely to

694 benefit from prospective monitoring and early immunotherapy interventions, thereby supporting

695 clinical decision-making and precision medicine.

696 **RESOURCE AVAILABILITY**

697 **Data availability statement**

698 The SILCODE dataset used in this study is not publicly available due to participant

699 privacy and ethical considerations. However, researchers interested in collaboration may contact

700   the corresponding authors to inquire about collaboration opportunities and apply for data access.

701   Applicants are required to provide a detailed research plan and sign a data use agreement. The

702   research team will respond within 30 business days upon receiving applications. All data sharing

703   will strictly adhere to relevant privacy protections and ethical guidelines to ensure the secure and

704   compliant use of the data.

705   **Code availability statement**

706   The code for the MRI-based deep learning model for Alzheimer's disease is available at

707   https://github.com/Chaogan-Yan/BrainImageNet. Our demonstration website

708   (http://brainimagenet.org) allows users to upload data directly and obtain risk scores after

709   providing informed consent. Additionally, the codes for statistical analyses, creating

710   individualized AD brain risk maps, and data visualization are available at

711   https://github.com/Chaogan-Yan/BrainImageNet/tree/master/ADPredictionPaperScripts.

712   **FUNDING AND ACKNOWLEDGEMENTS**

AUTHOR CONTRIBUTIONS

C-GY: Conceptualized and designed the research, interpreted results, revised the manuscript, and supervised the entire research process. YH: Conceptualized the research, interpreted results, acquired the data, and provided expert domain knowledge in the AD research field. JL: Interpreted results, acquired the data, and provided expert domain knowledge in the AD research field. BL: Conceptualized and designed the research, conducted data analysis and visualization, interpreted results, and drafted the manuscript. Y-RC: Designed the evaluation framework, conducted data analysis and visualization, interpreted results, and drafted the manuscript. R-XL, M-KZ, G-QC, and S-ZY: Acquired data and provided diagnostic and data collection details. FXC and PMT: Revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Livingston, G., J. Huntley, A. Sommerlad et al. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *The Lancet*. **396**:413–446. DOI:https://doi.org/10.1016/S0140-6736(20)30367-6.

2. Mintun, M.A., A.C. Lo, C. Duggan Evans et al. (2021). Donanemab in Early Alzheimer's Disease. *N Engl J Med*. **384**:1691–1704. DOI:https://doi.org/10.1056/NEJMoa2100708.

3. Van Dyck, C.H., C.J. Swanson, P. Aisen et al. (2023). Lecanemab in Early Alzheimer's Disease. *N Engl J Med*. **388**:9–21. DOI:https://doi.org/10.1056/NEJMoa2212948.

4. Hansson, O. (2021). Biomarkers for neurodegenerative diseases. *Nat Med*. **27**:954–963. DOI:https://doi.org/10.1038/s41591-021-01382-x.

5. Jia, J., Y. Ning, M. Chen et al. (2024). Biomarker Changes during 20 Years Preceding Alzheimer's Disease. *N Engl J Med*. **390**:712–722. DOI:https://doi.org/10.1056/NEJMoa2310168.

6. Leocadi, M., E. Canu, D. Calderaro et al. (2020). An update on magnetic resonance imaging markers in AD. *Ther Adv Neurol Disord*. **13**:175628642094798. DOI:https://doi.org/10.1177/1756286420947986.

7. Apostolova, L.G. (2023). The role of Neuroimaging in Patient Selection and Safety Monitoring during Lecanemab Treatment. *Alzheimer's & Dementia*. **19**:e077209. DOI:https://doi.org/10.1002/alz.077209.

755   8. Rathore, S., M. Habes, M.A. Iftikhar et al. (2017). A review on neuroimaging-based

756   classification studies and associated feature extraction methods for Alzheimer's disease and

757   its prodromal stages. *NeuroImage*. **155**:530–548.

758   DOI:https://doi.org/10.1016/j.neuroimage.2017.03.057.

759   9. Ebrahimighahnavieh, M.A., S. Luo and R. Chiong (2020). Deep learning to detect

760   Alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods*

761   *and Programs in Biomedicine*. **187**:105242.

762   DOI:https://doi.org/10.1016/j.cmpb.2019.105242.

763   10. Jo, T., K. Nho and A.J. Saykin (2019). Deep Learning in Alzheimer's Disease: Diagnostic

764   Classification and Prognostic Prediction Using Neuroimaging Data. *Front. Aging Neurosci.*

765   **11**:220. DOI:https://doi.org/10.3389/fnagi.2019.00220.

766   11. Martin, S.A., F.J. Townend, F. Barkhof et al. (2023). Interpretable machine learning for

767   dementia: A systematic review. *Alzheimer's & Dementia*. **19**:2135–2149.

768   DOI:https://doi.org/10.1002/alz.12948.

769   12. Qiu, S., P.S. Joshi, M.I. Miller et al. (2020). Development and validation of an interpretable

770   deep learning framework for Alzheimer's disease classification. *Brain*. **143**:1920–1933.

771   DOI:https://doi.org/10.1093/brain/awaa137.

772   13. Lu, B., X. Chen, F. Xavier Castellanos et al. (2024). The power of many brains: Catalyzing

773   neuropsychiatric discovery through open neuroimaging data and large-scale collaboration.

774   *Science Bulletin*. **69**:1536–1555. DOI:https://doi.org/10.1016/j.scib.2024.03.006.

775   14. Rathore, S., M. Habes, M.A. Iftikhar et al. (2017). A review on neuroimaging-based

776   classification studies and associated feature extraction methods for Alzheimer's disease and

777   its prodromal stages. *NeuroImage*. **155**:530–548.

778   DOI:https://doi.org/10.1016/j.neuroimage.2017.03.057.

779   15. Jo, T., K. Nho and A.J. Saykin (2019). Deep Learning in Alzheimer's Disease: Diagnostic

780   Classification and Prognostic Prediction Using Neuroimaging Data. *Front. Aging Neurosci.*

781   **11**. DOI:https://doi.org/10.3389/fnagi.2019.00220.

782   16. Ebrahimighahnavieh, M.A., S. Luo and R. Chiong (2020). Deep learning to detect

783   Alzheimer's disease from neuroimaging: A systematic literature review. *Computer Methods*

784   *and Programs in Biomedicine*. **187**:105242.

785   DOI:https://doi.org/10.1016/j.cmpb.2019.105242.

786   17. Lu, B., H.-X. Li, Z.-K. Chang et al. (2022). A practical Alzheimer's disease classifier via

787   brain imaging-based deep learning on 85,721 samples. *J Big Data*. **9**:101.

788   DOI:https://doi.org/10.1186/s40537-022-00650-y.

789   18. Barthélemy, N.R., G. Salvadó, S.E. Schindler et al. (2024). Highly accurate blood test for

790   Alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests. *Nat Med*.

791   **30**:1085–1095. DOI:https://doi.org/10.1038/s41591-024-02869-z.

792   19. Chohan, P., M. Dashwood, G. Theodoulou et al. (2022). Blood-based biomarkers for

793   Alzheimer's disease. *Prog Neurol Psychiatry*. **26**:10–14.

794   DOI:https://doi.org/10.1002/pnp.764.

795   20. Jack, C.R., J.S. Andrews, T.G. Beach et al. (2024). Revised criteria for diagnosis and staging

796   of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*.

797   **20**:5143–5169. DOI:https://doi.org/10.1002/alz.13859.

798   21. Li, X., X. Wang, L. Su et al. (2019). Sino Longitudinal Study on Cognitive Decline

799   (SILCODE): protocol for a Chinese longitudinal observational study to develop risk

800   prediction models of conversion to mild cognitive impairment in individuals with subjective

801   cognitive decline. *BMJ Open*. **9**:e028188. DOI:https://doi.org/10.1136/bmjopen-2018-

802   028188.

803   22. Li, X., X. Wang, L. Su et al. (2019). Sino Longitudinal Study on Cognitive Decline

804   (SILCODE): Protocol for a Chinese longitudinal observational study to develop risk

805   prediction models of conversion to mild cognitive impairment in individuals with subjective

806   cognitive decline. *BMJ Open*. **9**:e028188. DOI:https://doi.org/10.1136/bmjopen-2018-

807   028188.

808   23. McKhann, G.M., D.S. Knopman, H. Chertkow et al. (2011). The diagnosis of dementia due

809   to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's

810   Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's &*

811   *Dementia*. **7**:263–269. DOI:https://doi.org/10.1016/j.jalz.2011.03.005.

812   24. Morris, J.C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules.

813   *Neurology*. **43**:2412. DOI:https://doi.org/10.1212/WNL.43.11.2412-a.

814   25. Bondi, M.W., E.C. Edmonds, A.J. Jak et al. (2014). Neuropsychological Criteria for Mild

815   Cognitive Impairment Improves Diagnostic Precision, Biomarker Associations, and

816   Progression Rates. *J Alzheimers Dis*. **42**:275–289. DOI:https://doi.org/10.3233/JAD-140276.

817   26. Jessen, F., R.E. Amariglio, R.F. Buckley et al. (2020). The characterisation of subjective

818   cognitive decline. *The Lancet Neurology*. **19**:271–278. DOI:https://doi.org/10.1016/S1474-

819   4422(19)30368-0.

820   27. Folstein, M.F., S.E. Folstein and P.R. McHugh (1975). "Mini-mental state": A practical

821   method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric*

822   *Research*. **12**:189–198. DOI:https://doi.org/10.1016/0022-3956(75)90026-6.

823   28. Qh, G., Z. B, Z. Qh et al. (2012). Memory and Executive Screening (MES): A brief cognitive

824   test for detecting mild cognitive impairment. *BMC neurology*. **12**.

825   DOI:https://doi.org/10.1186/1471-2377-12-119.

826   29. Chen, K.-L., Y. Xu, A.-Q. Chu et al. (2016). Validation of the Chinese version of Montreal

827   Cognitive Assessment Basic for screening mild cognitive impairment. *J Am Geriatr Soc*.

828   **64**:e285–e290. DOI:https://doi.org/10.1111/jgs.14530.

829   30. Zhao, Q., Y. Lv, Y. Zhou et al. (2012). Short-term delayed recall of auditory verbal learning

830   test is equivalent to long-term delayed recall for identifying amnestic mild cognitive

831   impairment. *PLoS One*. **7**:e51157. DOI:https://doi.org/10.1371/journal.pone.0051157.

832   31. Guo Q., Jin L. and Hong Z. (2002). A specific phenomenon of Animal Fluency Test in

833   Chinese elderly. *Chinese Mental Health Journal*.

834   32. Guo Q.-H., Hong Z., Shi W.-X. et al. (1991). Boston Naming Test in Chinese elderly, patient

835   with mild cognitive impairment and Alzheimer's dementia. *Chinese Mental Health Journal*.

836   33. Zhao, Q., Q. Guo, F. Li et al. (2013). The Shape Trail Test: Application of a new variant of

837   the Trail Making Test. *PLoS One*. **8**:e57333.

838   DOI:https://doi.org/10.1371/journal.pone.0057333.

839   34. Lu, B., H.-X. Li, Z.-K. Chang et al. (2022). A practical Alzheimer's disease classifier via

840   brain imaging-based deep learning on 85,721 samples. *J Big Data*. **9**:101.

841   DOI:https://doi.org/10.1186/s40537-022-00650-y.

842   35. Janelidze, S., N. Mattsson, S. Palmqvist et al. (2020). Plasma P-tau181 in Alzheimer's

843   disease: Relationship to other biomarkers, differential diagnosis, neuropathology and

844   longitudinal progression to Alzheimer's dementia. *Nat Med*. **26**:379–386.

845   DOI:https://doi.org/10.1038/s41591-020-0755-1.

846   36. Carayol, J., C. Chabert, A. Di Cara et al. (2017). Protein quantitative trait locus study in

847   obesity during weight-loss identifies a leptin regulator. *Nat Commun*. **8**:2084.

848   DOI:https://doi.org/10.1038/s41467-017-02182-z.

849   37. R Core Team (2024). R: The R Project for Statistical Computing. https://www.r-project.org/.

850   38. Robin, X., N. Turck, A. Hainard et al. (2011). pROC: An open-source package for R and S+

851   to analyze and compare ROC curves. *BMC Bioinformatics*. **12**:77.

852   DOI:https://doi.org/10.1186/1471-2105-12-77.

853   39. Bates, D., M. Mächler, B. Bolker et al. (2015). Fitting Linear Mixed-Effects Models Using

854   lme4. *Journal of Statistical Software*. **67**:1–48. DOI:https://doi.org/10.18637/jss.v067.i01.

855   40. Kuznetsova, A., P.B. Brockhoff and R.H.B. Christensen (2017). lmerTest Package: Tests in

856   Linear Mixed Effects Models. *Journal of Statistical Software*. **82**:1–26.

857   DOI:https://doi.org/10.18637/jss.v082.i13.

858   41. Therneau, T.M. (2024). survival: Survival Analysis. Version 3.7-0.

859   42. Kassambara, A., M. Kosinski, P. Biecek et al. (2021). survminer: Drawing Survival Curves

860   using "ggplot2." Version 0.4.9.

861   43. Lenth, R.V. (2024). emmeans: Estimated Marginal Means, aka Least-Squares Means.

862   Version 1.10.4.

863   44. Jack, C.R., A. Arani, B.J. Borowski et al. (2024). Overview of ADNI MRI. *Alzheimers

864   Dement*. **20**:7350–7360. DOI:https://doi.org/10.1002/alz.14166.

865   45. Filippi, M., G. Cecchetti and F. Agosta (2023). MRI in the new era of antiamyloid mAbs for

866   the treatment of Alzheimer's disease. *Current Opinion in Neurology*. **36**:239–244.

867   DOI:https://doi.org/10.1097/WCO.0000000000001177.

868   46. Alzheimer's Disease Neuroimaging Initiative, A. Casamitjana, P. Petrone et al. (2018). MRI-

869   Based Screening of Preclinical Alzheimer's Disease for Prevention Clinical Trials. *JAD*.

870   **64**:1099–1112. DOI:https://doi.org/10.3233/JAD-180299.

871   47. Alzheimer's Association (2024). 2024 Alzheimer's disease facts and figures. *Alzheimers

872   Dement*. **20**:3708–3821. DOI:https://doi.org/10.1002/alz.13809.

873   48. Tang, Y., L. Zhao, Y. Lou et al. (2018). Brain structure differences between Chinese and

874   Caucasian cohorts: A comprehensive morphometry study. *Hum Brain Mapp*. **39**:2147–2155.

875   DOI:https://doi.org/10.1002/hbm.23994.

876   49. Chee, M.W.L., H. Zheng, J.O.S. Goh et al. (2011). Brain Structure in Young and Old East

877   Asians and Westerners: Comparisons of Structural Volume and Cortical Thickness. *Journal*

878   *of Cognitive Neuroscience*. **23**:1065–1079. DOI:https://doi.org/10.1162/jocn.2010.21513.

879   50. Tondelli, M., G.K. Wilcock, P. Nichelli et al. (2012). Structural MRI changes detectable up

880   to ten years before clinical Alzheimer's disease. *Neurobiology of Aging*. **33**:825.e25-825.e36.

881   DOI:https://doi.org/10.1016/j.neurobiolaging.2011.05.018.

882   51. Hirni, D.I., S.L. Kivisaari, S. Krumm et al. (2016). Neuropsychological Markers of Medial

883   Perirhinal and Entorhinal Cortex Functioning are Impaired Twelve Years Preceding

884   Diagnosis of Alzheimer's Dementia. *Journal of Alzheimer's Disease*. **52**:573–580.

885   DOI:https://doi.org/10.3233/JAD-150158.

886   52. van der Flier, W.M. and B.M. Tijms (2023). Treatments for AD: towards the right target at

887   the right time. *Nat Rev Neurol*. **19**:581–582. DOI:https://doi.org/10.1038/s41582-023-00869-

888   0.

889   53. Sims, J.R., J.A. Zimmer, C.D. Evans et al. (2023). Donanemab in early symptomatic

890   Alzheimer disease: The TRAILBLAZER-ALZ 2 randomized clinical trial. *JAMA*. **330**:512–

891   527. DOI:https://doi.org/10.1001/jama.2023.13239.

892   54. Simic, G., G. Stanic, M. Mladinov et al. (2009). Does Alzheimer's disease begin in the

893   brainstem? *Neuropathology Appl Neurobio*. **35**:532–554.

894   DOI:https://doi.org/10.1111/j.1365-2990.2009.01038.x.

895   55. Lee, J.H., J. Ryan, C. Andreescu et al. (2015). Brainstem morphological changes in

896   Alzheimer's disease. *NeuroReport*. **26**:411–415.

897   DOI:https://doi.org/10.1097/WNR.0000000000000362.

898   56. Weldemichael, D.A. and G.T. Grossberg (2010). Circadian Rhythm Disturbances in Patients

899   with Alzheimer's Disease: A Review. *International Journal of Alzheimer's Disease*. **2010**:1–

900   9. DOI:https://doi.org/10.4061/2010/716453.

901   57. Baloyannis, S.J., V. Costa and D. Michmizos (2004). Mitochondrial alterations Alzheimer's

902   disease. *Am J Alzheimers Dis Other Demen*. **19**:89–93.

903   DOI:https://doi.org/10.1177/153331750401900205.

904   58. Matsuoka, T., A. Imai, H. Fujimoto et al. (2018). Reduced Pineal Volume in Alzheimer

905   Disease: A Retrospective Cross-sectional MR Imaging Study. *Radiology*. **286**:239–248.

906   DOI:https://doi.org/10.1148/radiol.2017170188.

907   59. Liao, Y.-Z., J. Ma and J.-Z. Dou (2022). The Role of TDP-43 in Neurodegenerative Disease.

908   *Mol Neurobiol*. **59**:4223–4241. DOI:https://doi.org/10.1007/s12035-022-02847-x.

909   60. Josephs, K.A., M.E. Murray, J.L. Whitwell et al. (2016). Updated TDP-43 in Alzheimer's

910   disease staging scheme. *Acta Neuropathol*. **131**:571–585.

911   DOI:https://doi.org/10.1007/s00401-016-1537-1.

61. Guo, Y., X.-N. Shen, H.-F. Wang et al. (2023). The dynamics of plasma biomarkers across the Alzheimer's continuum. *Alz Res Therapy*. **15**:31. DOI:https://doi.org/10.1186/s13195-023-01174-0.

62. Ashton, N.J., W.S. Brum, G. Di Molfetta et al. (2024). Diagnostic accuracy of a plasma phosphorylated tau 217 immunoassay for Alzheimer disease pathology. *JAMA Neurol*. **81**:255. DOI:https://doi.org/10.1001/jamaneurol.2023.5319.

63. Janelidze, S., D. Bali, N.J. Ashton et al. (2023). Head-to-head comparison of 10 plasma phospho-tau assays in prodromal Alzheimer's disease. *Brain*. **146**:1592–1601. DOI:https://doi.org/10.1093/brain/awac333.

64. Hu, H., K.-L. Chen, Y.-N. Ou et al. (2019). Neurofilament light chain plasma concentration predicts neurodegeneration and clinical progression in nondemented elderly adults. *Aging*. **11**:6904–6914. DOI:https://doi.org/10.18632/aging.102220.

65. Yu, X., R. Shi, X. Zhou et al. (2024). Correlations between plasma markers and brain Aβ deposition across the AD continuum: Evidence from SILCODE. *Alzheimers Dement*. DOI:https://doi.org/10.1002/alz.14084.

66. Feinkohl, I., C.G. Schipke, J. Kruppa et al. Plasma Amyloid Concentration in Alzheimer's Disease: Performance of a High-Throughput Amyloid Assay in Distinguishing Alzheimer's Disease Cases from Controls. *J Alzheimers Dis*. **74**:1285–1294. DOI:https://doi.org/10.3233/JAD-200046.

67. Janelidze, S., C.E. Teunissen, H. Zetterberg et al. (2021). Head-to-Head Comparison of 8 Plasma Amyloid-β 42/40 Assays in Alzheimer Disease. *JAMA Neurol*. **78**:1375. DOI:https://doi.org/10.1001/jamaneurol.2021.3180.

68. Hampel, H., A. Elhage, M. Cho et al. (2023). Amyloid-related imaging abnormalities (ARIA): radiological, biological and clinical characteristics. *Brain*. **146**:4414–4424. DOI:https://doi.org/10.1093/brain/awad188.

69. Feng, X., F.A. Provenzano, S.A. Small et al. (2022). A deep learning MRI approach outperforms other biomarkers of prodromal Alzheimer's disease. *Alzheimer's Research & Therapy*. **14**:45. DOI:https://doi.org/10.1186/s13195-022-00985-x.

70. Popuri, K., D. Ma, L. Wang et al. (2020). Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. *Hum Brain Mapp*. **41**:4127–4147. DOI:https://doi.org/10.1002/hbm.25115.