A case-control clinical trial on the diagnostic performance for Alzheimer's Disease of a deep learning-based classification system using brain magnetic resonance imaging

Jong Bin Bae
  Seoul National University Bundang Hospital

Subin Lee
  Seoul National University

Hyunwoo Oh
  Vuno Inc

Jinkyeong Sung
  Vuno Inc

Dongsoo Lee
  Vuno Inc

Jiwon Han
  Seoul National University Bundang Hospital

Jun Sung Kim
  Seoul National University

Jae Hyoung Kim
  Seoul National University Bundang Hospital

Sang Eun Kim
  Seoul National University Bundang Hospital

Ki Woong Kim (✉ kwkimmd@snu.ac.kr)
  Seoul National University Bundang Hospital; Seoul National University College of Natural Sciences;
Seoul National University College of Medicine   https://orcid.org/0000-0002-1103-3858

Research

Abstract

Objective

To investigate diagnostic performance of a deep learning-based classification system using structural brain MRI (DLCS) for Alzheimer's disease (AD).

Methods

A single-center, case-control clinical trial was conducted. T1-weighted brain MRI scans of 188 patients with mild cognitive impairment or dementia due to AD and 162 cognitively normal controls were retrospectively collected. The patients were amyloid beta (Aβ)-positive, whereas the controls were Aβ-negative, on 18F-florbetaben positron emission tomography. Sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve were calculated to evaluate the performance of DLCS in the classification of Aβ-positive AD patients from Aβ-negative controls.

Results

The DLCS was excellent in classifying AD patients from normal controls; sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve for AD were 85.6% (95%CI, 79.8−90), 90.1% (95%CI, 84.5−94.2), 91.0% (95%CI, 86.3−94.1), 84.4% (95%CI, 79.2−88.5), and 0.937 (95%CI, 0.911−0.963), respectively.

Conclusion

The DLCS shows promise in clinical settings where it may improve early detection of AD in any individual who has undergone an MRI scan regardless of purpose.

Trial registration:

Korean Clinical Trials Registry, KCT0004758. Registered 21 February 2020, https://cris.nih.go.kr/cris/search/detailSearch.do/17665.

Introduction

The number of individuals with dementia is increasing globally. More than 130 million people are expected to live with dementia in 2050,[1] with Alzheimer's disease (AD) as the most prevalent type.[2] Since no cure for AD has been developed yet, early diagnosis is crucial for proper management of AD.

However, more than 60% of community-dwelling people with dementia are undiagnosed, due to its insidious nature [3]. To improve the accuracy and advance the timing of AD diagnosis, the National Institute on Aging-Alzheimer's Association proposed a new diagnostic criteria for AD that incorporated neuroimaging biomarkers such as amyloid beta (Aβ) deposition and neuronal degeneration [4–7]. However, while Aβ deposition is an earlier and more specific biomarker of AD than neurodegeneration, assessment of the former (i.e., positron emission tomography [PET]) has many practical drawbacks compared to the that of the latter (i.e., magnetic resonance imaging [MRI]), because PET scans are more expensive, involve radiation, and are less available in clinical settings.

Brain MRI is an effective and widely used tool for detecting neuronal loss and structural changes in the brain. Recently, several studies have developed artificial intelligence (AI)-based algorithms for classifying AD using structural brain MRI, with promising performance, including processing time and classification accuracy [8–13]. However, in most previous studies [9, 10, 12, 13], training and validation datasets were constructed by randomly splitting a dataset into two. Because the training and validation datasets came from the same population, the performance of algorithms were likely to be overestimated in those studies. Furthermore, most studies [9–13] did not confirm the presence of Aβ deposition in the AD patients as well as the absence of Aβ deposition in the normal controls, despite the fact that about 12% of the clinically diagnosed probable AD patients are Aβ-negative [14] and 10–40% of cognitively normal controls are Aβ-positive [15].

In our previous work, we developed a deep learning-based classification system for AD using structural brain MRI (DLCS) as an AI software as a medical device (SaMD) and found its accuracy in classifying probable AD patients from cognitively normal controls to be excellent (0.88–0.94) [16]. However, our previous work shared the same limitations as previous studies stated earlier. In addition, our previous work did not include mild cognitive impairment (MCI) due to AD in the patient group, which might have exaggerated the performance of the DLCS.

Therefore, in the current clinical trial, we investigated the performance of the DLCS in discriminating Aβ-positive patients with MCI or AD dementia from Aβ-negative cognitively normal controls, all of whom were from a sample independent of the population used for the development of the DLCS.

Methods

Study participants

A single-center, case-controlled clinical trial was conducted and registered in the Korean Clinical Trials Registry (KCT0004758). Data of subjects over 50 years of age who visited Seoul National University Bundang Hospital (SNUBH) and underwent a T1-weighted MRI scan between January 2010 and September 2019 were retrospectively collected. Our data include brain MRI scans with clinical assessment and 18F-florbetaben PET scans from visitors to our dementia clinic as well as from participants of the Korean Longitudinal Study on Cognitive Aging and Dementia (KLOSCAD) [17].

A group of patients with AD and a group with normal cognition (NC) matched for age and sex were screened and enrolled using the following inclusion criteria. The AD groups included those who had: (1) a diagnosis of probable or possible AD according to the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria, or MCI according to the International Working Group on MCI, and (2) amyloid deposition as determined by a positive 18F-florbetaben PET scan. The NC group included those who (1) had no subjective cognitive complaints, (2) had no objective cognitive decline in the Korean version of the Consortium to Establish a Registry for AD (CERAD-K) neuropsychological assessment battery, (3) were functioning independently in the community, and (4) had no amyloid deposition as determined by a negative 18F-florbetaben PET scan. Subjects who had any of the following conditions were excluded: (1) diagnosis of dementia with a cause other than or in addition to AD, i.e., mixed dementia, (2) brain pathologies on T1-weighted MRI that may cause cognitive deficits, (3) more than 1 year between the date of clinical assessment and date of MRI scan (NC and MCI participants only), and (4) white matter hyperintensities with a Fazeka's rating of 3 or higher on fluid-attenuated inversion recovery images.

The data of the participants were retrospectively screened and collected starting from April 27, 2020 to June 5, 2020 (6 weeks). The employment of the DLCS on the data were conducted between June 8, 2020 to June 19, 2020 (2 weeks).

Sample size calculation

We employed both the sensitivity and specificity of DLCS to AD as primary outcome measures. We calculated the sample size needed to evaluate whether DLCS performed better than a reference, based on a one-sided α of 2.5% (Z_α = 1.96), statistical power of 80% ($Z_{1-\beta}$ = 0.842), and the results of a pilot study. The pilot study tested the performance of DLCS using a dataset consisting of 367 AD patients and 316 controls with NC: 130 AD and 130 NC from SNUBH and 237 AD and 186 NC from the Alzheimer's Disease Neuroimaging Initiative database. At a threshold value of 0.38, the DLCS yielded a sensitivity of 82.0% (95% confidence interval [CI], 77.7−85.8%) and specificity of 83.2% (95%CI, 78.6−87.2). To calculate the sample size n, we used the following formula [18]:

$$
n = \frac{\left(Z_\alpha\sqrt{p_0(1-p_0)} + Z_{1-\beta}\sqrt{p_1(1-p_1)}\right)^2}{(p_1 - p_0)^2}
$$

where p_0 is the assumed sensitivity/specificity under the null hypothesis H_0, and p_1 is the targeted sensitivity/specificity under alternative hypothesis H_1. The p_0 and p_1 values were defined as the lower and higher bounds of the 95%CI of the sensitivity and specificity from the pilot study (p_0 = 0.777 and p_1 = 0.858 for sensitivity; p_0 = 0.786 and p_1 = 0.872 for specificity). The null hypothesis was that the sensitivity/specificity of the DLCS is less than or equal to the lower boundary of the assumed sensitivity/specificity. The alternative hypothesis was that it is higher. Based on this, the necessary

number of subjects with the disease was 188, and the number of subjects without the disease was 162. Therefore, the final estimated sample size was 350 subjects, consisting of 188 patients with AD and 162 normal controls that were matched for age (<5 years apart) and sex to the AD group.

Image acquisition

We acquired three-dimensional (3D) T1-weighted MR images in Digital Imaging and Communications in Medicine format using Philips Achieva and Ingenia scanners (Philips Medical Systems, Eindhoven, the Netherlands). The parameters were as follows: voxel dimensions = $1.0 \times 0.5 \times 0.5$ mm^3, slice thickness = 1.0 mm, echo time = 8.15 or 8.20 ms (for Achieva and Ingenia, respectively), repetition time = 4.61 ms, flip angle = 8°, and field of view = 240 × 240 mm.

We acquired 18F-florbetaben PET scans in 3D using a Discovery VCT scanner (General Electric Medical Systems, Milwaukee, WI, USA). The subjects were injected with 8.1 mCi (300 MBq) 18F-florbetaben (Neuraceq) through a slow single intravenous bolus (6 MBq) in a total volume of 10 mL. After a 90-min uptake period, 20-min PET images comprising four 5-min dynamic frames were obtained. Images of each time frame were reframed into one summed frame. Board-certified nuclear medicine physicians then determined Aβ-positivity based on visual interpretation of tracer uptake in the gray matter compared to neighboring subcortical white matter in the following four brain regions: the temporal lobes, frontal lobes, posterior cingulate cortex/precuneus, and parietal lobes.

Deep learning-based Alzheimer's disease classification system

We used VUNO Med-DeepBrain AD (version 1.0.0, VUNO Inc., Seoul, South Korea), which is the DLCS for AD. The convolutional neural network model used in VUNO Med-DeepBrain AD has been previously described [16]. Briefly, the DLCS receives a subject's T1-weighted image, extracts coronal slices from areas that span the medial temporal lobe, and feeds each coronal slice as a separate input into a convolutional neural network. The network, which uses Inception-V4 as its backbone, extracts various features that include structural and textural information of the brain from the coronal slice. The feature vector is then concatenated with the subject's age and sex information (which is input to the system at the beginning with the MRI scan) and the location information (slice number) of the coronal slice, and entered into a fully connected network that calculates the probability of the slice belonging to that of a patient with AD. The probabilities of each slice are averaged to calculate a final score that represents the subject's probability of having AD (score ranges from 0 to 1).

In this clinical trial, we processed the MRI data of subjects anonymously, omitting information that could identify the individual (name, sex, birth date, and hospital number). A researcher (K.J.S.), who was blinded to the subjects' clinical diagnoses and did not participate in the construction of the study dataset, performed the processing of the subjects' data with DLCS. The DLCS was installed on a desktop PC with the following specifications: Intel hexa-core 2.90 GHz CPU with 16 GB RAM running on Ubuntu 18.04.4 LTS.

Statistical analysis

We evaluated the accuracy of the DLCS in the diagnosis of AD by comparing its output (a continuous probability ranging from 0 to 1) with the subject's clinical diagnosis. We defined sensitivity and specificity as the primary outcomes, and the area under the receiver operating characteristic curve (AUC) as the secondary outcome.

Continuous variables were compared using independent samples t-test, and categorical variables were compared using the chi-square test between groups. We estimated the 95%CIs of sensitivity and specificity using the Clopper-Pearson method [19] and the AUC using the DeLong test [18]. All statistical analyses were performed using SPSS, version 20 (SPSS Inc., Chicago, IL, USA) and MedCalc (version 16.4.3; MedCalc Software, Mariakerke, Belgium).

Standard Protocol Approvals, Registrations, and Patient Consents

This clinical trial (Korean Clinical Trials Registry identifier: KCT0004758) was approved by the Ministry of Food and Drug Safety in South Korea and the Institutional Review Board of SNUBH. The design and conduct of this study were in accordance with the principles outlined in the Declaration of Helsinki [20]. Because this clinical trial was conducted retrospectively, participation consent forms from subjects or legal guardians of the subjects were waived.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Results

We enrolled a total of 350 subjects who met the eligibility criteria, with 162 (46.3%) in the NC group and 188 (53.7%) in the AD group. The demographic and clinical characteristics of the participants are summarized in Table 1. The mean age of the whole dataset was 73.3 ± 7.23 (range, 55 to 92) years. Age and sex were comparable between the NC and AD groups, while years of education were higher in the NC group. In the patient group, 76 (40.4%) had MCI due to AD, and the rest had AD dementia. All participants with MCI due to AD had a clinical dementia rating (CDR) score of 0.5. Among the 112 participants with AD dementia, 68 (60.7%) had a CDR score of 0.5, 35 (31.3%) had a CDR score of 1, and the rest (8.0%) had a CDR score of 2 or 3. The models of MR scanners were comparable, while the type of head coil was different between the two groups.

Table 1
Subject characteristics

	NC (N = 162)	AD (N = 188)	*t* or *χ2*	*P* Value
Age (years), mean ± SD	73.3 ± 6.9	73.9 ± 7.4	−0.8	0.42
Age (years), No. (%)				
50−59	0 (0.0%)	12 (6.4%)		
60−69	46 (28.4%)	36 (19.1%)		
70−79	84 (51.9%)	96 (51.1%)		
80−89	32 (19.7%)	43 (22.9%)		
≥90	0 (0.0%)	1 (0.5%)		
Sex, No. (%)			0.001	0.97
Male	54 (33.4%)	63 (33.5%)		
Female	108 (66.6%)	125 (66.5%)		
Education (years), mean ± SD	12.4 ± 4.5	11.1 ± 4.9	2.57	0.01
MMSE (score), mean ± SD	27.5 ± 2.2	20.9 ± 4.9	16.21	< 0.001
MRI				
Scanner, No. (%)			1.64	0.44
Philips Achieva	137 (84.6%)	167 (88.8%)		
Philips Ingenia	20 (12.3%)	18 (9.6%)		
Philips Ingenia CX	5 (3.1%)	3 (1.6%)		
Head coil, No. (%)			63.79	< 0.001
SENSE-Head-8	73 (45.1%)	39 (20.7%)		
SENSE-NV-16	15 (9.3%)	89 (47.4%)		
Dual coil	42 (25.9%)	39 (20.7%)		
Multi coil	32 (19.7%)	21 (11.2%)		

Abbreviations: NC = normal control; AD = Alzheimer's disease; MMSE = Mini-Mental State Examination

As summarized in Fig. 1, DLCS demonstrated a good diagnostic performance for AD. Its sensitivity for AD was 85.6% (95%CI, 79.8−90.3), and the lower bound of 95%CIs for its sensitivity was higher than the assumed value of 77.7%. Its specificity for AD was 90.1% (95%CI, 84.5−94.2), and the lower bound of

95%CIs for its specificity was higher than the assumed value of 78.6%. Its accuracy, positive predictive value, and negative predictive value for AD were 87.7%, 91.0%, and 84.4%, respectively. The AUC of DLCS for AD classification was 0.937 (95%CI, 0.911−0.963).

Discussion

This clinical trial demonstrated that the diagnostic performance for AD of DLCS was excellent. To the best of our knowledge, this is the first clinical trial in the field of AI-based AD diagnosis using structural brain MRI data.

According to a previous meta-analysis, the rate of undetected dementia in community-dwelling elderly is pooled to be 61.7% (95%CI = 55.0−68.0%) [3]. The rate of undetected cases is especially higher for those with dementia that have a slow progressive onset, such as AD [21].

Structural brain MRI has been extensively explored for improving the early diagnosis of AD because of its good accessibility and rich information on neurodegeneration. Various automated MRI measures that are sensitive to AD detection, such as cortical volume [22], cortical thickness [23], shape [24], and texture [25] have been developed. The previously reported MRI-based markers reported good accuracy in diagnosing AD, but they all require heavy preprocessing of data and a long time to process, which is not feasible in a non-research setting such as in the clinic. This critical limitation has long delayed the active use of MRI-based markers in clinical practice. However, in our previous work, DLCS was found to take only 23 s per case to process structural brain MR images. DLCS could drastically reduce the processing time by making an inference using a prelearned neural network, rather than preprocessing the whole data and extracting or calculating new features [16]. In addition, the features extracted by DLCS can comprehensively reflect volumetric, shape, and textural information, making it potentially more informative than previously developed single MRI measure-markers. What remained now for employing DLCS in clinical practice was to prove its diagnostic performance.

In this clinical trial, the diagnostic performance of the DLCS for AD was found to be excellent according to the criteria of excellent biomarkers proposed by the Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group on "Molecular and Biochemical Markers of Alzheimer's Disease" [26]. The working group suggested that an excellent evaluating biomarker should have a sensitivity approaching or exceeding 85%, a specificity of approximately 75−85% or greater, and a positive predictive value of approximately 80% or more. The sensitivity, specificity, and positive predictive value of the DLCS were 85.6%, 90.1%, and 91.0%, respectively, which met the requirements proposed by the working group.

The diagnostic performance of the DLCS for AD was comparable to or better than that of clinical diagnosis, fluorodeoxyglucose PET, and cerebral blood flow single-photon emission computed tomography (SPECT). Clinical diagnosis of probable AD according to the NINCDS-ADRADA criteria has

shown a sensitivity of 70.9–75.3% and a positive predictive value of 59.5–70.8% for autopsy-proven AD. [27] Fluorodeoxyglucose PET has a sensitivity of 84% and specificity of 74% for autopsy-proven AD [28] and a sensitivity of 75.8% and specificity of 74.3% for amyloid PET-proven AD.[29] Cerebral blood flow SPECT showed a sensitivity of 63% and specificity of 82% to autopsy-proven AD [30] and sensitivity of 42.9% and specificity of 82.9% to amyloid PET-proven AD.[29]

Both amyloid PET[31] and cerebrospinal fluid (CSF) β-amyloid 42[32] can detect AD much earlier during the preclinical stage than structural brain MRI. However, in clinical settings, they are not supposed to be administered to patients with AD during the preclinical stage.[33] In contrast, structural brain MRI is conducted not only for diagnosing various types of dementia but also for diagnosing other numerous neurologic disorders and even for health checkups in clinical settings. Since the DLCS can be applied to all structural brain MRI scans taken for any purpose, it can increase the detection rate of AD, which may otherwise go unnoticed, and direct the patients for a timely examination that confirms the presence of AD.

This study has several strengths. First, we minimized the misclassification bias by confirming the presence of Aβ in the AD group and the absence of Aβ in the NC group using 18F-florbetaben PET scans. Assigning patients to the AD group based solely on clinical diagnosis can result in a misenrollment of patients who have AD-like symptoms but do not actually have AD pathology. Likewise, using only clinical diagnosis can also result in misassigning asymptomatic AD patients to the NC group, as it is known that up to one-fourth of cognitively normal elderly individuals can have Aβ pathology, which is recognized as a preclinical form of AD.[15] Second, we minimized the overestimation of the diagnostic performance of the DLCS by including MCI due to AD in the patient group [5]. Since neurodegenerative changes in the brain are less prominent in the prodromal phase, the diagnostic performance of the biomarkers using structural brain MRI can be overestimated if the patients with prodromal AD (MCI) would not be included in the patient group.

Limitations

This study has several limitations. First, all MRI scans used in this study were acquired from a single scanner (Philips) using the same protocol. Therefore, the performance of the DLCS on scans from other vendors or protocols is unknown. Second, DLCS currently only takes 3D T1-weighted images as input data because 3D scans contain higher anatomical detail and resolution than conventional 2D scans. However, 3D scans are not always available in clinical settings, which may undermine the applicability of DLCS. Third, it is not clear which features contributed to the predictions made by the DLCS, which can undermine the explainability of the results. Increasing the explainability and interpretability of deep learning algorithms will be crucial in increasing the trustworthiness of the technology for use in the medical domain. This is an unresolved issue that is currently the topic of many recent research.[34].

Conclusions

In conclusion, DLCS, a software as a medical device using structural brain MRI, demonstrated excellent diagnostic performance for AD. When used together during screening of MRI, taken for whatever purpose, DLCS may improve the early detection of AD.

Abbreviations

AD: Alzheimer's disease

DLCS: Deep learning-based classification system

MCI: Mild cognitive impairment

MRI: Magnetic resonance imaging

NC: Normal cognition

PET: Positron emission tomography

SNUBH: Seoul National University Bundang Hospital

Declarations

Ethics approval and consent to participate

The protocol for the current study was approved by the Ministry of Food and Drug Safety in South Korea and the Institutional Review Board of SNUBH.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing Interests:

J.B. Bae, S. Lee, J.S. Kim, J.W. Han, and K.W. Kim received royalty income from VUNO Inc.

H. Oh, J. Sung, and D. Lee are employee of VUNO Inc.

The other authors declare no conflicts of interest.

Funding

Authors' Contributions

J.B. Bae and S. Lee contributed equally to this work and should be considered co-first authors. J.B. Bae and S. Lee had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: J.B. Bae, S. Lee, H. Oh, J. Sung, K.W. Kim

Acquisition, analysis, or interpretation of data: J.B. Bae, S. Lee

Drafting of manuscript: J.B. Bae, S. Lee, H., K.W. Kim

Critical revision of the manuscript for intellectual content: All authors

Statistical analysis: J.B. Bae, S. Lee

Obtained funding: K.W. Kim

Administrative, technical, and material support: J.S. Kim, J.W. Han

Study supervision: K.W. Kim.

Acknowledgements

Not applicable.

Authors' information

Jong Bin Bae: jongbinbae@hotmail.com; Subin Lee: subin9126@hotmail.com; Hyunwoo Oh: zero92311@gmail.com; Jinkyeong Sung: bebop8231@gmail.com; Dongsoo Lee: ehdtnsla@gmail.com; Ji Won Han: jwhanmd@snu.ac.kr; Jun Sung Kim: 019kli@hanmail.net; Jae Hyoung Kim: jaehkim@snu.ac.kr; Sang Eun Kim: kse@snu.ac.kr; Ki Woong Kim: kwkimmd@snu.ac.kr

References

1. Prince MJ. World Alzheimer Report 2015: the global impact of dementia: an analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International; 2015.
2. Kim KW, Park JH, Kim M-H, Kim MD, Kim B-J, Kim S-K, Kim JL, Moon SW, Bae JN, Woo JI. A nationwide survey on the prevalence of dementia and mild cognitive impairment in South Korea. J Alzheimers Dis. 2011;23:281–91.

3. Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, Danat IM, Zhou W, Copeland JR, Anstey KJ, Chen R. Prevalence and determinants of undetected dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017;7:e011146.

4. Jack CR Jr, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, Thies B, Phelps CH. Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:257–62.

5. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:270–9.

6. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:280–92.

7. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:263–9.

8. Li H, Habes M, Wolk DA, Fan Y, Alzheimer's Disease Neuroimaging I, the Australian Imaging B. Lifestyle Study of A: **A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data**. Alzheimers Dement. 2019;15:1059–70.

9. Li F, Liu M. Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. Comput Med Imaging Graph. 2018;70:101–10.

10. Luo S, Li X, Li J. Automatic Alzheimer's Disease Recognition from MRI Data Using Deep Learning Method. Journal of Applied Mathematics Physics. 2017;05:1892–8.

11. Liu M, Zhang J, Adeli E, Shen D. Landmark-based deep multi-instance learning for brain disease diagnosis. Med Image Anal. 2018;43:157–68.

12. Basaia S, Agosta F, Wagner L, Canu E, Magnani G, Santangelo R, Filippi M. Alzheimer's Disease Neuroimaging I: **Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks**. Neuroimage Clin. 2019;21:101645.

13. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF. Alzheimer's Disease Neuroimaging I: **Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images**. Sci Rep. 2018;8:5697.

14. Ossenkoppele R, Jansen WJ, Rabinovici GD, Knol DL, van der Flier WM, van Berckel BNM, Scheltens P, Visser PJ, Group atAPS. Prevalence of Amyloid PET Positivity in Dementia Syndromes: A Meta-analysis. JAMA. 2015;313:1939–50.

15. Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FR, Visser PJ, Amyloid Biomarker Study G, Aalten P, Aarsland D, et al. Prevalence of cerebral amyloid pathology in persons

without dementia: a meta-analysis. JAMA. 2015;313:1924–38.

16. Bae JB, Lee S, Jung W, Park S, Kim W, Oh H, Han JW, Kim GE, Kim JS, Kim JH, Kim KW. Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Sci Rep. 2020;10:22252.

17. Han JW, Kim TH, Kwak KP, Kim K, Kim BJ, Kim SG, Kim JL, Kim TH, Moon SW, Park JY, et al. Overview of the Korean Longitudinal Study on Cognitive Aging and Dementia. Psychiatry Investig. 2018;15:767–74.

18. DeLong ER, DeLong DM, Clarke-Pearson DL. **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach**. *Biometrics* 1988:837–845.

19. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 1934;26:404–13.

20. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Jama. 2013;310:2191–4.

21. Sternberg SA, Wolfson C, Baumgarten M. Undetected dementia in community-dwelling older people: the Canadian Study of Health and Aging. J Am Geriatr Soc. 2000;48:1430–4.

22. Fox NC, Warrington EK, Freeborough PA, Hartikainen P, Kennedy AM, Stevens JM, Rossor MN. Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. Brain. 1996;119(Pt 6):2001–7.

23. Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN, Grodstein F, Wright CI, Blacker D, Rosas HD, et al. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. Cereb Cortex. 2009;19:497–510.

24. Gerardin E, Chételat G, Chupin M, Cuingnet R, Desgranges B, Kim HS, Niethammer M, Dubois B, Lehéricy S, Garnero L, et al. Multidimensional classification of hippocampal shape features discriminates Alzheimer's disease and mild cognitive impairment from normal aging. Neuroimage. 2009;47:1476–86.

25. Lee S, Lee H, Kim KW. Magnetic resonance imaging texture predicts progression to dementia due to Alzheimer disease earlier than hippocampal volume. J Psychiatry Neurosci. 2020;45:7–14.

26. The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group: **Consensus report of the Working Group on**: "**Molecular and Biochemical Markers of Alzheimer's Disease**".. In *Neurobiol Aging*, vol. 19. 1998/04/29 edition. pp. 109–116; 1998:109–116.

27. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. J Neuropathol Exp Neurol. 2012;71:266–73.

28. Jagust W, Reed B, Mungas D, Ellis W, Decarli C. **What does fluorodeoxyglucose PET imaging add to a clinical diagnosis of dementia?** *Neurology* 2007, 69:871–877.

29. Nadebaum DP, Krishnadas N, Poon AM, Kalff V, Lichtenstein M, Villemagne VL, Jones G, Rowe CC. **A head-to-head comparison of cerebral blood flow SPECT and (18) F-FDG PET in the diagnosis of Alzheimer's Disease**. *Intern Med J* 2020.

30. Jagust W, Thisted R, Devous MD, Sr., Van Heertum R, Mayberg H, Jobst K, Smith AD, Borys N. SPECT perfusion imaging in the diagnosis of Alzheimer's disease: a clinical-pathologic study. Neurology. 2001;56:950–6.

31. La Joie R, Ayakta N, Seeley WW, Borys E, Boxer AL, DeCarli C, Doré V, Grinberg LT, Huang E, Hwang JH, et al. Multisite study of the relationships between antemortem [(11)C]PIB-PET Centiloid values and postmortem measures of Alzheimer's disease neuropathology. Alzheimers Dement. 2019;15:205–16.

32. Maddalena A, Papassotiropoulos A, Müller-Tillmanns B, Jung HH, Hegi T, Nitsch RM, Hock C. Biochemical diagnosis of Alzheimer disease by measuring the cerebrospinal fluid ratio of phosphorylated tau protein to beta-amyloid peptide42. Arch Neurol. 2003;60:1202–6.

33. Laforce R, Rosa-Neto P, Soucy JP, Rabinovici GD, Dubois B, Gauthier S. Canadian Consensus Guidelines on Use of Amyloid Imaging in Canada: Update and Future Directions from the Specialized Task Force on Amyloid imaging in Canada. Can J Neurol Sci. 2016;43:503–12.

34. Tjoa E, Guan C. **A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI**. *IEEE Trans Neural Netw Learn Syst* 2020, **Pp**.

Figures
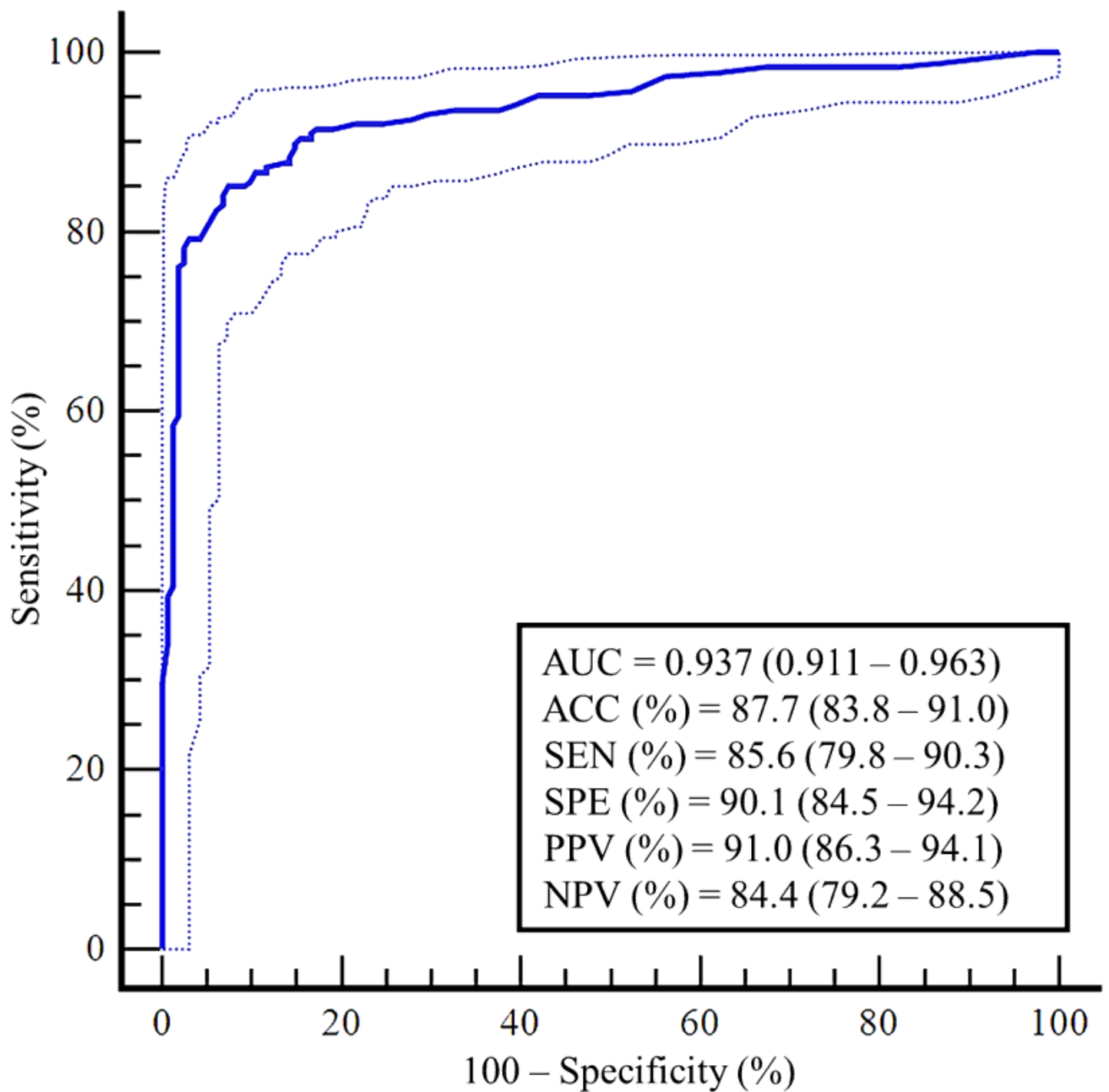
Figure 1

Receiver operating characteristic curve of Alzheimer's disease discrimination The receiver operating characteristic curve and performance metrics are shown with 95% confidence intervals in parentheses. Abbreviations: AUC = area under the curve; ACC = accuracy; SEN = sensitivity; SPE = specificity; PPV = positive predictive value; NPV = negative predictive value