Machine learning of brain structural biomarkers for Alzheimer's disease (AD) diagnosis, prediction of disease progression, and amyloid beta deposition in the Japanese population

著者	SHIINO Akihiko, SHIRAKASHI Yoshitomo, ISHIDA Manabu, TANIGAKI Kenji
journal or publication title	Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring
volume	13
number	1
year	2021-10-14
URL	http://hdl.handle.net/10422/00013376

doi: 10.1002/dad2.12246(https://doi.org/10.1002/dad2.12246)

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations aremade.

Received: 6 June 2021 | Accepted: 11 June 2021 | Published online: 11 October 2021

DOI: 10.1002/dad2.12246

Alzheimer's & Dementia
Diagnosis, Assessment & Disease Monitoring

RESEARCH ARTICLE

Machine learning of brain structural biomarkers for Alzheimer's disease (AD) diagnosis, prediction of disease progression, and amyloid beta deposition in the Japanese population

Akihiko Shiino[1] | **Yoshitomo Shirakashi**[1] | **Manabu Ishida**[3] | **Kenji Tanigaki**[2] | **Japanese Alzheimer's Disease Neuroimaging Initiative**

[1] Molecular Neuroscience Research Center, Shiga University of Medical Science, Shiga, Japan

[2] Research Institute, Shiga Medical Center, Shiga, Japan

[3] Department of Neurology, Shimane University, Shimane, Japan

Correspondence
Akihiko Shiino, Molecular Neuroscience Research Center, Shiga University of Medical Science, Seta, Ohtsu, Shiga 520-2192, Japan.
Email: shiino@belle.shiga-med.ac.jp

Abstract

Introduction: We developed machine learning (ML) designed to analyze structural brain magnetic resonance imaging (MRI), and trained it on the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. In this study, we verified its utility in the Japanese population.

Methods: A total of 535 participants were enrolled from the Japanese ADNI database, including 148 AD, 152 normal, and 235 mild cognitive impairment (MCI). Probability of AD was expressed as AD likelihood scores (ADLS).

Results: The accuracy of AD diagnosis was 88.0% to 91.2%. The accuracy of predicting the disease progression in non-dementia participants over a 3-year observation was 76.0% to 79.3%. More than 90% of the participants with low ADLS did not progress to AD within 3 years. In the amyloid positron emission tomography (PET)–positive MCI, the hazard ratio of progression was 2.39 with low ADLS, and 5.77 with high ADLS. When high ADLS was defined as N+ and Pittsburgh compound B (PiB) PET positivity was defined as A+, the time to disease progression for 50% of MCI participants was 23.7 months in A+N+, whereas it was 52.3 months in A+N-.

Conclusion: These results support the feasibility of our ML for the diagnosis of AD and prediction of the disease progression.

KEYWORDS
ADNI, Alzheimer's disease, artificial intelligence, machine learning, MRI

1 | INTRODUCTION

The increased age of the global population has led to an increase in the prevalence of dementia. Alzheimer's Disease International estimated that >50 million people developed dementia worldwide in 2019; in addition, a new case of dementia is diagnosed every 3 seconds.[1] In 2014, the social cost of dementia in Japan was ≈14.5 trillion yen; it is estimated to reach 24.3 trillion yen by 2060.[2] The rapid growth of

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* published by Wiley Periodicals, LLC on behalf of Alzheimer's Association

artificial intelligence (AI) has brought significant changes in the industrial structure, including the field of medical imaging. The introduction of AI into routine diagnostic imaging is expected to improve diagnostic accuracy and reduce differences among clinicians, leading to reduced medical costs. Machine learning (ML), the core algorithm of AI, requires a large amount of reliable data, and large data such as Alzheimer's Disease Neuroimaging Initiative (ADNI) studies are most suitable for constructing AI for analyzing brain structure through magnetic resonance imaging (MRI).[3,4]

Jo et al.[5] systematically reviewed articles and selected the top 16 studies in which ML was used for AD diagnosis and/or predicting mild cognitive impairment (MCI) progression by fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET) and MRI. Although these studies have shown high performance of ML, it should be noted that they lack validation based on independent trial data and do not take into account metrics to maximize the overall accuracy of the classification model while maintaining generalizability. We recently developed ML that was trained in the North American ADNI, and validated in the Australian ADNI database. Our ML performed well even in the validation database compared to the previous reports. Furthermore, when its performance was compared with two expert neuroradiologists, our ML outperformed the radiologists for AD diagnosis even though they were provided with information about the degree of the hippocampal atrophy.[6]

The use of ML has also been applied to predict amyloid beta (Aβ) deposition before the onset of the disease. The prodromal phase of AD can last for decades, and early detection of the pathophysiological changes might be useful for preventive intervention with disease-modifying therapies (DMTs). Biomarker measurement from cerebrospinal fluid (CSF) or PET has been established to detect the presence of AD pathology, but they are neither affordable nor widely available in all clinical settings because of their invasiveness and high cost. Brain MRI is non-invasive and has been used widely to rule out other abnormalities in subjects with suspected AD. The application of ML to the analysis of structural MRI may have the potential to predict the presence of AD pathology in the prodromal stage. There are several studies showing that MRI combined with ML can predict amyloid positivity with sufficient accuracy to be cost-effective as a pre-screening tool.[7–9] With this background, it is necessary to examine whether our ML could predict cerebral Aβ deposition.

The purpose of this study was to investigate the applicability of our ML-based algorithm to the Japanese population. We examined the relationship between the results of the algorithm and the risk of disease progression, biomarkers from CSF, and imaging biomarkers from amyloid PET. We evaluated the effectiveness of the algorithm in the diagnosis of AD, the prediction of disease progression in individuals without dementia, and the risk estimation of brain Aβ deposition. The present study is the first empirical report of the application of the ML to the Japanese population. Traditionally, hippocampal atrophy has been used as an important biomarker for AD on MRI, probably because it can be easily assessed visually. In this study, maximizing the extraction of features from the brain structure over the entire brain enables prediction of Aβ accumulation and high risk for disease progression. As a result, it is considered to be useful for the preliminary selection of DMT subjects that will be needed in the future.

RESEARCH IN CONTEXT

1. **Systematic Review**: The authors reviewed the literature using traditional sources (such as PubMed) and the "progress reports" of the Alzheimer's Disease Neuroimaging Initiative (ADNI) published yearly in *Alzheimer's & Dementia*.
2. **Interpretation**: We showed that our machine learning (ML) algorithm, which is trained on the North American ADNI database, performed well in the Japanese population. Hippocampal atrophy is an important and easily identifiable biomarker of neuronal injury in AD visualized using routine magnetic resonance imaging (MRI). However, mild hippocampal atrophy can occur pathologically even in Braak's stage IV or earlier; the sole reliance on the presence of hippocampal atrophy can limit the diagnosis of AD and prediction of disease progression. ML can analyze a large amount of information and guide to an optimal conclusion. This means extracting biomarkers from MRI that cannot be achieved by conventional methods. Our experience in adjusting the hyperparameters that are manually set and managing variables to introduce into ML have led to excellent results in the Japanese population.
3. **Future Direction**: The availability of disease-modifying therapy against amyloid beta (Aβ) has increased in clinical practice. The issue lies in formulating an efficient strategy for candidate selection. A consensus has been reached regarding the benefits of antihypertensive agents in the prevention of stroke. However, this consensus is due to the ease of monitoring blood pressure in daily practice. Because cerebral amyloidosis is often observed in normal elderly people, we believe that technology is also needed to screen for the risk of progression to AD by using widely available MRI. The specificity of the method may be improved by using biomarkers from blood and other sources in the future.

2 | METHODS

2.1 | Japanese ADNI database

The Japanese ADNI (JADNI) database was used to obtain data from 535 participants consisting of 148 patients with AD, 235 with MCI, and 152 normal individuals (NL). In 2007, the JADNI was initiated as a longitudinal, multi-center cohort study of healthy elderly adults and patients with MCI and early AD. The current study included native

Japanese speakers between the ages of 60 and 84, who were accompanied by a study partner. Approximately 25% of the participants underwent 1.5T and 3.0T MR examinations, but only 1.5T images were used in this study. Pittsburgh compound B (PiB) enhanced PET scan was performed on a total of 156 participants, including 41 AD, 53 NL, and 62 MCI participants. At baseline, CSF biomarkers were measured on a total of 297 participants including 54 AD, 54 NL, and 189 MCI participants.

The JADNI data were extracted from the National Bioscience Database Center (hum0043.v1). The study protocol (UMIN000001374) was approved by the institutional review committees at the 38 participating clinical sites, and the ethics committee of our university. Informed written consent was obtained from all participants at each clinical site.

2.2 | MRI Analyses

All MR images at baseline were downloaded and analyzed using software named brain anatomical analysis using diffeomorphic deformation (BAAD, version 4.4) that was developed in our laboratory. More information about BAAD can be found elsewhere.[6] We inducted anatomical region of interest (ROI) as a unit to reduce dimensionality, and created an algorithm with a support vector machine (SVM). Technically, the introduction of standardized variables that have been adjusted for covariates improves ML performance. Voxel-based morphometry (VBM) is an algorithm capable of standardizing variables by unifying the brain shape through coordinate transformations and correcting the local volume with covariates for each voxel. Details of a standard VBM procedure were discussed in a previous study.[10] Because neural connections form a functional or anatomical unit in the brain, hypertrophy or atrophy of neurons appears as clusters, which is recognized as "autocorrelation" among voxels. This seems to be true in AD atrophy as well, because tau lesions may spread from cell to cell via neuronal connections.[11] For comparison, we used software named Voxel-based Specific Regional Analysis System for Alzheimer's Disease (VSRAD) as a representative example. This software simply examines the degree of atrophy of medial temporal structures. The details of VSRAD were discussed in previous studies.[12,13] The VSRAD software has an ROI in the medial temporal structures where atrophy is common in AD patients including the entorhinal cortex, hippocampus, and amygdala.

We used the radial basis function as the kernel for the SVM, and the values of parameters were optimized using the North American ADNI database. The BAAD software expressed the probability of an AD diagnosis as an AD likelihood score (or ADLS). ADLS represents the distance to the hyperplane and is obtained from the posterior probability function $Pr = (Y = k|X = x)$, the probability is the class Y is the class k given that the input variable X is x. The probability is transformed by a sigmoid function to compress the value within the range of [0, 1]; the larger the value, the more likely is the diagnosis of AD. In addition, we used the Mini Mental State Examination (MMSE) score for the ML and expressed the probability as ADLS cognitive (ADLSc).

2.3 | Amyloid beta imaging

PET imaging was performed with three-dimensional (3D) dynamic scans and intravenous administration of PiB. All PET images taken at each site were subjected to the J-ADNI PET quality control (QC) process, which corrects head motion between frames before creating a total frame image.[14]

The positive and negative results were based on the JADNI PiB PET central reading judgment. A positive result of Aβ deposition was considered when the PiB accumulation was found in any region of the four cortices spreading over more than one brain gyrus, and was evidently higher than the white matter found in the precuneus/posterior cingulate gyri, frontal, lateral temporal, and lateral parietal lobes. The definition of positive, equivocal, and negative regional uptake has been described in a previous study.[15]

2.4 | CSF biomarker measurement

CSF samples were examined for the presence of A$\beta_{1\text{-}42}$, total tau (tTau), and phosphorylated tau (pTau) using Innogenetics (INNO-BIA AlzBio3; Ghent, Belgium) and the multiplex xMAP Luminex platform (Luminex Corp., Austin, TX, USA). The results were validated at the J-ADNI Biomarker Core at Niigata University.[16] The optimal cutoff values for A$\beta_{1\text{-}42}$, pTau, and tTau were obtained using receiver-operating characteristic (ROC) curves for discrimination of the individuals with AD (n = 54) and NL (n = 54) at baseline.

2.5 | Statistics

Statistical analyses were performed using JMP software (version 14.3, SAS Institute, Cary, NC, USA). ROC curves were used to assess the ability of the model to classify diseases. The optimal cutoff point was determined using the Youden index method. Details of the equations for accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1, and Matthews correlation coefficient (MCC) have been described in previous studies.[17,18] The significance level was set at $P < .05$.

3 | RESULTS

Among the individuals who underwent PiB PET (n = 156), three were clinically diagnosed with AD but with negative PiB accumulation. Therefore, we set a subgroup consisting of 38 AD, 62 MCI, and 53 NL, excluding these 3 individuals. The demographic features of the individuals are summarized in Table 1. The AD and MCI groups were significantly older than the NL group (t-test, $P < .001$), and the MMSE and clinical dementia rating values were significantly different among the three groups (Wilcoxon, $P < .001$).

TABLE 1 Demographic features of the Japanese ADNI

Group	NL(*)		MCI([a])		AD([b])	
Subjects(n)	152	53	235	62	148	38
Age	68.2 ± 5.6[a,b]	66.4 ± 4.6[a,b]	73.3 ± 5.8[*]	72.0 ± 5.4[*b]	73.9 ± 6.6[*]	74.3 ± 6.7[*b]
Sex (F/M)	80/72	26/27	117/118	31/31	85/63	22/16
Education (y)	13.8 ± 2.8[a,b]	13.8 ± 2.4[b]	13.0 ± 2.8[*]	13.7 ± 2.9	12.4 ± 3.2[*]	12.7 ± 2.7[*]
MMSE	29.1 ± 1.2[a,b]	29.3 ± 1.1[a,b]	26.3 ± 1.8[*b]	26.6 ± 1.9[*b]	22.5 ± 1.8[*a]	22.2 ± 1.8[*a]
CDR memory	0.0 ± 0.0[a,b]	0.0 ± 0.0[a,b]	0.54 ± 0.15[*b]	0.55 ± 0.15[*b]	0.96 ± 0.37[*a]	0.95 ± 0.38[*a]
CDR global	0.0 ± 0.0[a,b]	0.0 ± 0.0[a,b]	0.50 ± 0.03[*b]	0.50 ± 0.00[*b]	0.67 ± 0.24[*a]	0.64 ± 0.23[*a]
CDT[§]	4.6 ± 1.0[a,b]	4.6 ± 1.1[a,b]	4.2 ± 1.1[*b]	4.3 ± 1.1[*]	3.9 ± 1.3[*a]	3.9 ± 1.4[*]
BNT	28.0 ± 4.9[a,b]	28.0 ± 5.8[a,b]	25.4 ± 5.1[*b]	26.7 ± 4.9[*b]	24.2 ± 5.2[*a]	24.5 ± 4.8[*a]
ADAS-11	4.6 ± 2.6[a,b]	4.6 ± 2.6[a,b]	10.9 ± 4.4[*b]	9.5 ± 4.5[*b]	16.1 ± 4.9[*a]	16.5 ± 4.2[*a]
ApoE4 number	0.25 ± 0.46[a,b]	0.36 ± 0.56[a,b]	0.59 ± 0.62[*b]	0.53 ± 0.62[*b]	0.76 ± 0.72[*a]	0.54 ± 0.64[*a]
ApoE4 (0/1/2)	116/34/2	36/15/2	114/104/17	33/25/4	61/62/25	19/14.26/3

Abbreviations: AD, Alzheimer's disease; ADAS-11, Alzheimer's Disease Assessment Scale-Cognitive with 11 tasks; ApoE, apolipoprotein E; BNT, Boston Naming Test; CDR, Clinical Dementia Rating; CDT, Clock Drawing Test; NL, cognitively normal subjects; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination.

Gray boxes are the data from the subjects who underwent PiB PET.

The special symbols in parentheses in the top row are symbols for displaying the significant difference between the two groups. For example, in the NL column. [a,b]Indicates that there is a statistically significant difference ($P < .05$) between NL-MCI and NL-AD.

Age and years of education were tested by *t*-test; the others were tested by Wilcoxon test.

[c]The full score for CDT is five points; approximately circular dial, symmetry of number placement, correctness of numbers, presence of two hands, and hands being oriented correctly.

TABLE 2 Performance of BAAD and VSRAD for AD/NL discrimination

	All participants (n = 300)			Participants with PiB PET (n = 91)		
	ADLS	ADLSc	VSRAD	ADLS	ADLSc	VSRAD
AUC	0.9461	0.9930	0.9178	0.9727	0.9995	0.9499
Accuracy (%)	88.0	95.3	85.0	91.2	97.8	86.8
Sensitivity (%)	85.1	95.3	84.5	89.5	100.0	86.8
Specificity (%)	90.8	95.4	85.5	92.5	96.2	86.8
PPV (%)	90.0	95.3	85.0	89.5	95.0	82.5
NPV (%)	86.3	95.4	85.0	92.5	100.0	90.2
F1 (%)	87.5	95.3	84.7	89.5	97.4	84.6
MCC (%)	76.1	90.7	70.0	81.9	95.6	73.2
PLR	9.2	20.7	5.8	11.9	26.5	6.6
NLR	0.2	0.0	0.2	0.1	0.0	0.2
Post-test odds	9.0	20.1	5.7	8.5	19.0	4.7

Abbreviations: AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive rate; MCC, Matthews correlation coefficient; PLR, positive likelihood ratio; NLR, negative likelihood ratio.

3.1 | Diagnostic performance for AD

The optimal cutoff values determined by the Youden index to classify AD and NL in the JADNI database were 0.42, 0.23, and 1.26 for ADLS, ADLSc, and VSRAD, respectively. Because the optimal cutoff value for ADLS validated using the North American ADNI database was 0.48,[6] the cutoff points for ADLS and ADLSc were set at 0.5, and for VSRAD at 1.3 for the subsequent analysis. Accuracy, sensitivity, and specificity are influenced by prevalence (eg., the percentage of AD and NL patients of the data); therefore, the MCC and F1 scores were also expressed. The ROC curves and statistical results are shown in Table 2. When including all participants, the accuracy was measured at 95.3% for ADLSc, followed by 88.0% and 85.0% for ADLS and VSRAD, respectively. Similarly, in the subgroup excluding PiB PET-negative AD, the

TABLE 3 Prediction of Aβ deposition in each algorithm

	AD (n = 41)			MCI (n = 62)			NL (n = 53)		
	ADLS	ADLSc	VSRAD	ADLS	ADLSc	VSRAD	ADLS	ADLSc	VSRAD
AUC	0.8333	0.8509	0.7105	0.8258	0.8316	0.7613	0.5606	0.5985	0.4634
Accuracy (%)	87.8	100.0	82.9	79.0	77.4	72.6	75.5	79.2	73.6
Sensitivity (%)	89.5	100.0	86.8	80.5	82.9	75.6	0.0	0.0	11.1
Specificity (%)	66.7	–	33.3	76.2	66.7	66.7	90.9	95.5	86.4
PPV (%)	97.1	100.0	94.3	86.8	82.9	81.6	0.0	0.0	14.3
NPV (%)	33.3	–	16.7	66.7	66.7	58.3	81.6	82.4	82.6
F1 (%)	93.2	100.0	90.4	83.5	82.9	78.5	0.0	0.0	12.5
MCC (%)	41.4	–	14.9	55.1	49.6	41.1	-12.9	-9.0	-2.8
PLR	2.7	–	1.3	3.4	2.5	2.3	0.0	0.0	0.8
NLR	0.2	–	0.4	0.3	0.3	0.4	1.1	1.0	1.0
Post-test odds	34.0	–	16.5	6.6	4.9	4.4	0.0	0.0	0.2

Abbreviations: AD, Alzheimer's disease; MCI, mild cognitive impairment; NL, normal subjects; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive rate; MCC, Matthews correlation coefficient; PLR, positive likelihood ratio; NLR, negative likelihood ratio.

accuracies of the ADLS, ADLSc, and VSRAD were 91.2%, 97.8%, and 86.8%, respectively. The positive likelihood ratio (PLR) was high for ADLS and ADLSc, ranging from 9.2 to 11.9 and 20.7 to 26.5, respectively, whereas it was 5.8 to 6.6 for VSRAD; these values indicate the effectiveness of the examination. Likelihood ratios above 10 and below 0.1 are thought to provide strong evidence to rule in or rule out diagnoses in most circumstances.[19] In particular, the extremely high discrimination ability of ADLS and ADLSc is supported by their likelihood ratios, which indicates the excellent performance of the algorithm implemented in the BAAD software.

3.2 | Prediction of amyloid beta deposition from ML-MRI

In the AD group, the PiB-positive rates were 97.1% and 100%, at the cutoffs of ADLS >0.5 and ADLSc >0.5, respectively. In the NL group, the PiB-positive rate was 0% for both ADLS >0.5 and ADLSc >0.5 cutoffs. In the MCI group, at the cutoffs of ADLS >0.5 and ADLSc >0.5, the PiB-positive rates were 86.8% and 82.9%, respectively (Table 3).

3.3 | Determination of the cutoff values for CSF biomarkers

The distributions of $A\beta_{1\text{-}42}$, pTau, and tTau concentrations in the CSF from 54 AD and 54 NL participants are shown in Figure S1. The optimal cutoff values of $A\beta_{1\text{-}42}$, pTau, tTau, and pTau/$A\beta_{1\text{-}42}$ ratio determined by the Youden index method were 333 pg/mL, 44 pg/mL, 86 pg/mL, and 0.14, respectively. The cutoff value of 333 pg/mL for $A\beta_{1\text{-}42}$ was equal to that in a previous report.[20] Because the mean age of the two groups were different (AD: 73.2 ± 5.9; NL: 68.3 ± 5.5), age was accounted for in the model.

3.4 | Prediction of disease progression and risk assessment by biomarkers

During the 3 years of observation, among the 387 participants with NL or MCI, 119 MCI had converted to AD, 4 NL to MCI, and one NL to AD. Ten cases of reverse conversion from MCI to NL and one case from AD to MCI were observed. A total of 111 participants underwent PiB PET, and 3 participants who converted to AD were PiB negative at the initial visit. The prediction accuracy of disease progression in non-dementia individuals was 77.4% to 79.4% for ADLS and 79.6% to 80.4% for ADLSc. The PPV suggested that 50% to 60% of non-dementia individuals with high (>0.5) ADLS or ADLSc converted to AD within 3 years. Conversely, >90% of the individuals with low (<0.5) ADLS or ADLSc did not progress to AD for 3 years. For reference, the prediction accuracy for disease progression by PiB PET was 75.7%. The details of the results can be seen in the Supplemental Table.

Kaplan-Meier plots for the conversion from MCI to AD in each biomarker are shown in Figure 1 (with 95% confidence intervals in Figure S2). In this figure, results from PiB PET and ADLS were shown as solid lines, and those from CSF $A\beta_{1\text{-}42}$ (Aβ) and pTau were shown as dotted lines. MCI individuals with amyloid PET positive A(+) and high ADLS N(+) or CSF Aβ positive Aβ (+) and pTau positive pTau(+) progressed faster than the MCI with amyloidosis, A(+)N(−) or Aβ(+)pTau(−). Similar to MCI participants with negative biomarkers, A(−)N(−) or Aβ(−)pTau(−), MCI with amyloid PET negative (A−) and high ADLS N(+) were less likely to convert to AD. In the MCI individuals who are amyloid PET positive, the proportional hazard ratio (HR) of conversion from MCI to AD was 2.39 when ADLS was low, whereas it was 5.77 when ADLS was high. The estimated time to conversion for 50% of the MCI participants with amyloid PET-positive and high ADLS A(+)N(+) was 23.7 months, for 90% was 62.5 months. The estimated time to conversion for 50% of the MCI participants with A(+)N(−)

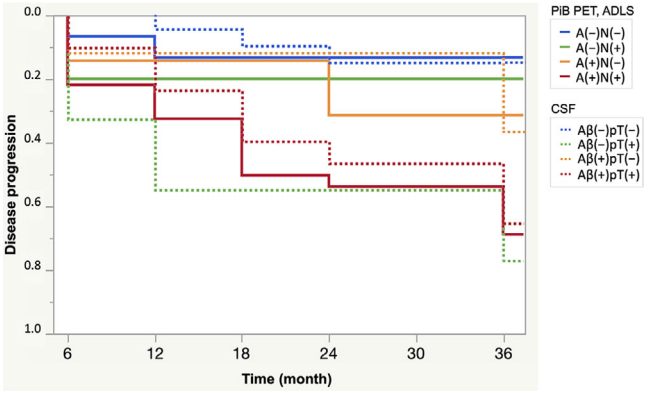

FIGURE 1 Progression rate from mild cognitive impairment (MCI) to Alzheimer's disease (AD). Kaplan-Meier analysis for disease progression from MCI to AD. In this figure, results from Pittsburgh compound B (PiB) positron emission tomography (PET) (A) and AD likelihood scores (ADLS) (N) are shown as solid lines, and those from cerebrospinal fluid (CSF) $A\beta_{1\text{-}42}$ (amyloid beta, Aβ) and phosphorylated tau (pT) are shown as dotted lines

was 52.3 months, and for 90% was 137.9 months. The HRs for each biomarker are summarized in Table 4.

4 | DISCUSSION

In this study, we investigated the usefulness of our VBM-ML algorithms that uses information from MMSE and MRI. MMSE is the widely accepted as a simple test to provide an overall measure of cognitive function in clinical practice, whereas MRI is the most available imaging modality to determine the cause of dementia, such as neurodegeneration, infarction, tumor, and hydrocephalus. Because ADNI is a simplified model that excludes dementia other than AD, it is premised that several examinations such as assessment of symptoms and signs, neuropsychological tests, dopamine transporter scans, metaiodobenzylguanidine (MIGB) scintigraphy, and single-photon emission computed tomography (SPECT) for cerebral blood flow are considered to exclude other types of dementia.

Hippocampal atrophy is an important biomarker for AD and is necessary in its investigation. VSRAD has been established in Japan as an adjunct tool for AD diagnosis by demonstrating the degree of atrophy (*z*-score) of medial temporal structures.[21] This software is used widely in Japan, and the Japanese Radiological Society (JRS) recommends its use. We investigated whether regions other than the medial temporal structures could be used for the diagnosis of AD, and tried to present the degree of atrophy in whole brain regions. However, the difficulty of analyzing numerous amounts of data has led us to introduce our ML algorithm for analysis. In this way, we expected that the algorithm could adapt to the varied atrophy patterns of AD. As a result, our ML outperformed VSRAD in diagnosing AD, predicting Aβ deposition and disease progression in non-dementia cases.

We estimated that the optimal cutoff of the *z*-score for VSRAD in JADNI was 1.3, which was lower compared to a previous report.[22] A standard deviation of 1.3 is considered to be in the normal range; therefore, it may be difficult for clinicians to visually determine hippocampal atrophy. The low cutoff value may result from the study design of the ADNI that targeted patients with mild dementia with an MMSE score of at least 20. Given that the hippocampal atrophy is

TABLE 4 Hazard ratio of MCI to AD conversion

Level 1	Level 2	HR	95% CI	Wald test
N(+)	N(−)	3.22	1.99-5.22	<0.0001*
A(+)N(−)	A−N−	2.39	0.33-17.08	0.3856
A(+)N(+)	A+N−	2.42	0.56-10.47	0.2388
A(+)N(+)	A−N−	5.77	1.31-25.35	0.0203*
A(−)N(+)	A−N−	1.84	0.16-21.17	0.6258
Aβ(+)pT(−)	Aβ(−)pT(−)	2.08	0.41-10.42	0.3751
Aβ(+)pT(+)	Aβ(+)pT(−)	2.26	0.68-7.48	0.1814
Aβ(+)pT(+)	Aβ(−)pT(−)	4.69	1.41-15.59	0.0116*
Aβ(−)pT(+)[a]	Aβ(−)pT(−)	8.38	2.04-34.39	0.0032*
Aβ(+)tT(-)	Aβ(-)tT(-)	5.62	1.15 - -27.43	0.0326*
Aβ(+)tT(+)	A(+)tT(-)	1.04	0.44-2.44	0.9355
Aβ(+)tT(+)	Aβ(-)tT(-)	5.83	1.36-25.06	0.0179*
Aβ(-)tT(+)[a]	Aβ(-)tT(-)	9.53	1.91-46.80	0.0055*

Abbreviations: A, amyloid deposition from Pittsburgh compound B (PiB) PET; N, neuronal injury from ADLS; Aβ (amyloid beta) , CSF $A\beta_{1\text{-}42}$; pT, CSF phosphorylated tau; tT, CSF total tau. The sign indicates positive (+) and negative (−).

HR: proportional hazard ratio over the 3-year follow-up adjusted for age, sex, and education period.

*$P < .05$ in Wald test.

[a]Of the 12 cases of Aβ(-)pTau(+), five underwent PiB PET, and three were amyloid positive. Similarly, of the 16 cases of Aβ(-)tTau(+), six underwent PiB PET, and four were amyloid positive.

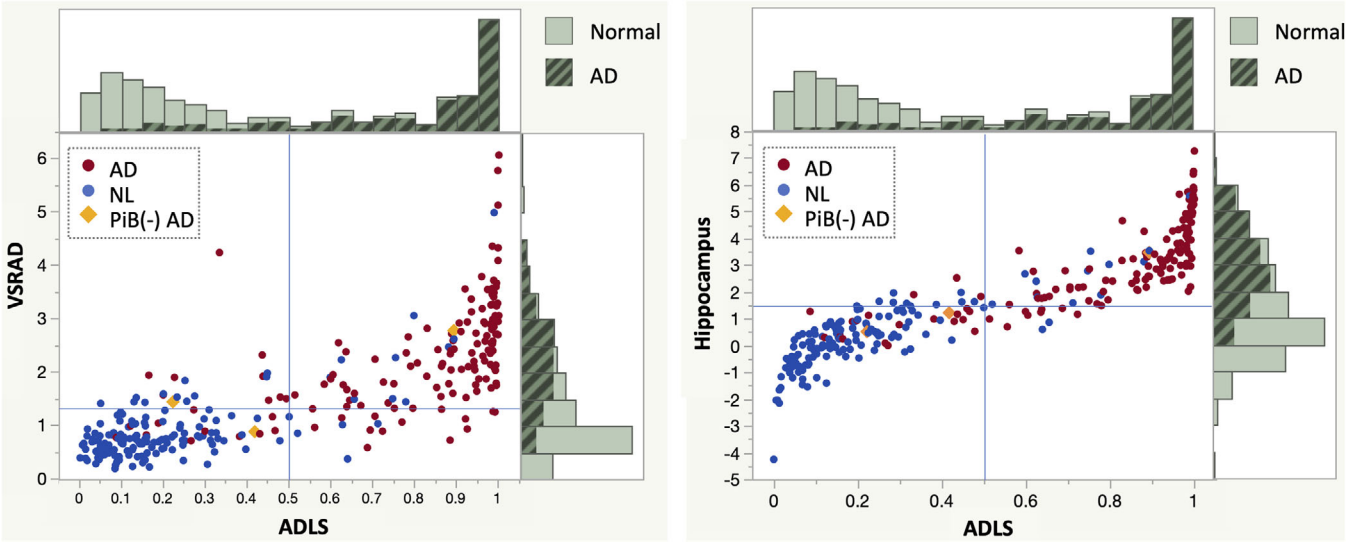

FIGURE 2 Scatter plot between AD likelihood scores (ADLS) and *z*-score of voxel-based specific regional analysis system for Alzheimer's disease (VSRAD) or hippocampus. Left panel: The horizontal axis represents the ADLS, and the vertical axis represents the *z*-score of VSRAD. The subjects in the lower right quadrant are subjects with less medial temporal atrophy, but with higher ADLS. The upper left quadrant is a group of subjects with medial temporal atrophy, but with lower ADLS. Right panel: The horizontal axis represents the ADLS, and the vertical axis represents the averaged z-score of the hippocampus on both sides. In the lower right quadrant, 12/16 patients (75.0%) were diagnosed with AD. In the upper left quadrant, 17/21 (81.0%) subjects were NL. AD, Alzheimer's disease; NL, cognitively normal subject; PiB(-)AD; subjects who were clinically diagnosed with AD but Pittsburgh compound B negative.

relatively identifiable visually, cortical atrophy may be even more difficult to assess visually in the early stages of AD. VSRAD supports radiologists by providing information of hippocampal atrophy, but our ML outperformed radiologists even when supported by VSRAD,[6] our ML also outperformed VSRAD in the Japanese population (Table 1). Of interest, the ML was also effective in predicting cerebral Aβ accumulation (Table 3). The hippocampal-sparing (HS) type is found in ≈15% of AD patients.[23] As shown in Figure 2, some AD patients with less pronounced hippocampal atrophy were also found in JADNI, although not strictly the "HS type" as defined by Murray et al.[24] It was notable that the BAAD ML suggested the possibility of AD even if the hippocampal atrophy was inconspicuous, and conversely, the possibility of NL even if the hippocampal atrophy was significant. The presence of neurofibrillary tangles (NFTs) in the neocortex, corresponding to stage III/IV or higher,[25] is important in establishing a pathological diagnosis of AD; this would indicate the need for an MRI evaluation of brain atrophy outside the medial temporal lobe. Although VSRAD focuses on hippocampal atrophy, BAAD ML seems to perform flexible diagnostic prediction by learning the atrophy pattern of the whole brain.

The selection of individuals for disease-modifying therapy (DMT) for Aβ deposition will be an important challenge when it becomes clinically applicable. In the prospective population-based Mayo Clinic Study of Aging in Olmsted County in Minnesota,[26] the prevalence of Aβ positivity was 2.7% in the 50 to 59 age group, 18.3% in the 60 to 69 age group, 32.1% in the 70 to 79 age group, and 41.2% in the 80 to 89 age group. In Aβ-positive individuals, the HR for progression to MCI in those without dementia was 2.26 times higher than that of Aβ-negative individuals; and the HR for progression to AD in Aβ-positive MCI was 1.86 times higher than that of Aβ-negative MCI. In our estimation with Cox proportional hazards model adjusting for age, sex, and education in JADNI, the HR for progression to AD in Aβ-positive MCI compared with Aβ-negative MCI was 2.39, which was similar to the results of the Mayo Clinic Study (Table 4). Accumulation of cerebral Aβ is a prominent risk factor for AD; however, one-third of the population over the age of 70 was Aβ positive, and the rate increased with age. Therefore, it is not practical to lead all elderly people to PET or CSF examinations without prior selection. MRI is readily feasible and less invasive, and our ML predicts cerebral Aβ deposition with high accuracy from MRI in non-dementia participants, which may minimize the amount of PET or CSF tests.

Another consideration is that the selection of patients who will undergo DMT depends on the cost-benefit performance of the therapy, so it may be unlikely that all Aβ-positive patients will be eligible for prophylactic DMT for Aβ removal from the brain. In our estimation, the HR for progression to AD in PiB-positive MCI with high ADLS (>0.5) was 5.77 compared with low ADLS (<0.5) MCI (Table 4). This result indicates that our ML can detect MCI patients, who are likely to develop dementia in the near future.

In the results of CSF biomarker measurement (Table 4), the Aβ-negative groups with abnormal pTau or tTau levels were more likely to progress to dementia, even greater than in the Aβ positive with pTau positive or tTau positive groups. At this time, there is no clear evidence to explain these results. It is known that amyloid PET and CSF biomarkers do not match in 10% to 20% of cases.[27] Of the 12 Aβ-negative, pTau positive MCI subjects, five underwent PiB PET, of which three were Aβ positive. Similarly, of the 16 Aβ-negative, tTau-positive MCI subjects, six underwent PiB PET, of which four were Aβ positive. This showed that there were not a few cases that were Aβ negative by CSF

measurement but Aβ positive by PiB PET. It is unclear whether this is due to the inherent nature of the examinations, the stage of the disease, or the speed of disease progression. At least, elevated tau protein in CSF reflects neuronal damage, which is consistent with disease progression.

As the pathogenesis of AD becomes clear, the NIA-AA research framework has proposed a new diagnostic approach called the ATN system,[28] which is based on biomarkers instead of the conventional clinical diagnosis. According to this system, a positive finding of amyloid in PET or decreased $A\beta_{1\text{-}42}$ concentration in CSF (A+) and increased phosphorylated tau in CSF or positive tau PET (T+), that is, (A+T+) is similar in concept to the diagnostic criteria for pathologic AD. The specific distribution of tau-PET signals is a strong indicator of the topography of future atrophy at the single patient level, and there was a strong relationship between baseline tau-PET and subsequent brain atrophy (N+).[29] Because AD pathology can be defined as A+T+ by biomarkers, the optimal timing of DMT for Aβ suppression seems to be before subsequent N+.

Considering this, it may be consequential to estimate the duration until conversion to AD when it becomes N+. In this study, the Kaplan-Meier analysis showed that the estimated time for 50% MCI to convert to AD was 26.5 months for patients with A+ on PiB PET and 23.7 months for patients with A+N+ on PiB PET and ADLS. Therefore, when these biomarkers are positive, half of MCI patients have only about 2 years before dementia; the investigation of the effectiveness of DMT in this short period remains an issue for future work. Recently, the 221AD301 ENGAGE study reported that the clearance of aggregated Aβ can reduce the cognitive decline due to AD pathology and, above all, that the DMT window is still open in patients with MCI or early-stage AD.[30]

This study has several limitations. First, NFT pathology is not the sole cause of brain atrophy, and the pattern analysis of brain atrophy alone limits specificity. In particular, TDP-43 is strongly associated with hippocampal atrophy in the elderly, which is a vexing problem in MRI analysis. In addition, 4R-tauopathy and cerebral amyloidosis are associated with cerebral atrophy; therefore, molecular-targeted PET is necessary in atypical cases. Second, as was often discussed in studies using the ADNI database, patients with dementia other than AD were carefully excluded from the model. Therefore, the clinical relevance of AD diagnosis in this study is limited to situations in which patients have been screened for other causes of dementia. Third, the definition of MCI in the ADNI study is biased toward the amnestic type rather than a clinical heterogenous MCI sample. The short observation period is a drawback for the prediction analysis in this study. Some non-dementia individuals may eventually progress to AD if the follow-up period is long enough. Therefore, the results of the study cannot be adapted for long-term preventive treatment.

In conclusion, this study supports the feasibility of our ML algorithm for the diagnosis of AD and prediction of disease progression in the Japanese population. By assessing the whole brain with VBM, the information obtained from MRI can be dramatically enhanced, and ML is effective for this information processing.

ACKNOWLEDGMENTS

J-ADNI was supported by the following grants: Translational Research Promotion Project from the New Energy and Industrial Technology Development Organization of Japan; Research on Dementia, Health Labor Sciences Research Grant; Life Science Database Integration Project of Japan Science and Technology Agency; Research Association of Biotechnology (contributed by Astellas Pharma Inc., Bristol-Myers Squibb, Daiichi-Sankyo, Eisai, Eli Lilly and Company, Merck-Banyu, Mitsubishi Tanabe Pharma, Pfizer Inc., Shionogi & Co., Ltd., Sumitomo Dainippon, and Takeda Pharmaceutical Company), Japan, and a grant from an anonymous foundation.

CONFLICTS OF INTEREST

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORCID

Akihiko Shiino https://orcid.org/0000-0001-6203-9339

REFERENCES

1. 2019 Alzheimer's disease facts and figures. *Alzheimer's Dement*. 2019;15:321-387.
2. Sado M, Ninomiya A, Shikimoto R, et al. The estimated cost of dementia in Japan, the most aged society in the world. *PLoS One*. 2018;13:e0206508.
3. Weiner MW, Veitch DP, Aisen PS, et al. Recent publications from the Alzheimer's disease neuroimaging initiative: reviewing progress toward improved AD clinical trials. *Alzheimers Dement*. 2017;13:e1-e85.
4. Choy G, Khalilzadeh O, Michalski M, et al. Current applications and future impact of machine learning in radiology. *Radiology*. 2018;288:318-328.
5. Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. *Front Aging Neurosci*. 2019;11:220.
6. Syaifullah AH, Shiino A, Kitahara H, Ito R, Ishida M, Tanigaki K. Machine learning for diagnosis of AD and prediction of MCI progression from brain MRI using brain anatomical analysis using diffeomorphic deformation. *Front Neurol*. 2021;11:576029.
7. Casamitjana A, Petrone P, Tucholka A, et al. MRI-based screening of preclinical Alzheimer's disease for prevention clinical trials. *J Alzheimers Dis*. 2018;64:1099-1112.
8. Kim JP, Kim J, Jang H, et al. Predicting amyloid positivity in patients with mild cognitive impairment using a radiomics approach. *Sci Rep*. 2021;11:6954.
9. Petrone PM, Casamitjana A, Falcon C, et al. Prediction of amyloid pathology in cognitively unimpaired individuals using voxel-wise analysis of longitudinal structural brain MRI. *Alzheimers Res Ther*. 2019;11:72.
10. Kurth F, Gaser C. Luders E. A 12-step user guide for analyzing voxel-wise gray matter asymmetries in statistical parametric mapping (SPM). *Nat Protoc*. 2015;10:293-304.
11. Vogel JW, Iturria-Medina Y, Strandberg OT, et al. Spread of pathological tau proteins through communicating neurons in human Alzheimer's disease. *Nat Commun*. 2020;11:2612.
12. Matsuda H, Mizumura S, Nemoto K, et al. Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated lie algebra improves the diagnosis of

probable Alzheimer disease. *AJNR Am J Neuroradiol*. 2012;33:1109-1114.
13. Matsuda H. MRI morphometry in Alzheimer's disease. *Ageing Res Rev*. 2016;30:17-24.
14. Ikari Y, Nishio T, Makishi Y, et al. Head motion evaluation and correction for PET scans with 18F-FDG in the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) multi-center study. *Ann Nucl Med*. 2012;26:535-544.
15. Yamane T, Ishii K, Sakata M, et al. Inter-rater variability of visual interpretation and comparison with quantitative evaluation of (11)C-PiB PET amyloid images of the Japanese Alzheimer's Disease Neuroimaging Initiative (J-ADNI) multicenter study. *Eur J Nucl Med Mol Imaging*. 2017;44:850-857.
16. Shaw LM, Vanderstichele H, Knapik-Czajka M, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol*. 2009;65:403-413.
17. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. 2006;27:861-874.
18. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim et Biophys Acta Protein Struct*. 1975;405:442-451.
19. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ*. 2004;329:168-169.
20. Iwatsubo T, Iwata A, Suzuki K, et al. Japanese and North American Alzheimer's disease neuroimaging initiative studies: harmonization for international trials. *Alzheimers Dement*. 2018;14:1077-1087.
21. Matsuda H. Voxel-based morphometry of brain MRI in normal aging and Alzheimer's disease. *Aging Dis*. 2013;4:29-37.
22. Sone D, Imabayashi E, Maikusa N, et al. Voxel-based Specific Regional Analysis System for Alzheimer's Disease (VSRAD) on 3-tesla normal database: diagnostic accuracy in two independent cohorts with early Alzheimer's disease. *Aging Dis*. 2018;9:755-760.
23. Ferreira D, Nordberg A, Westman E. Biological subtypes of Alzheimer disease: a systematic review and meta-analysis. *Neurology*. 2020;94:436-448.
24. Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol*. 2011;10:785-796.
25. Crary JF, Trojanowski JQ, Schneider JA, et al. Primary age-related tauopathy (PART): a common pathology associated with human aging. *Acta Neuropathol*. 2014;128:755-766.
26. Roberts RO, Aakre JA, Kremers WK, et al. Prevalence and outcomes of amyloid positivity among persons without dementia in a longitudinal, population-based setting. *JAMA Neurol*. 2018;75:970-979.
27. de Wilde A, Reimand J, Teunissen CE, et al. Discordant amyloid-beta PET and CSF biomarkers and its clinical consequences. *Alzheimers Res Ther*. 2019;11:78.
28. Jack CR, Jr., Bennett DA, Blennow K, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14:535-562.
29. La Joie R, Visani AV, Baker SL, et al. Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Sci Transl Med*. 2020;12(524):eaau5732.
30. Arnold C. Post-hoc analysis could give new life to the Alzheimer's drug aducanumab. *Nat Med*. 2020.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Shiino A, Shirakashi Y, Ishida M, Tanigaki K. Machine learning of brain structural biomarkers for Alzheimer's disease (AD) diagnosis, prediction of disease progression, and amyloid beta deposition in the Japanese population. *Alzheimer's Dement*. 2021;13:e12246. https://doi.org/10.1002/dad2.12246