Received: 25 July 2022 | Revised: 24 May 2023 | Accepted: 9 June 2023

DOI: 10.1002/hbm.26407

RESEARCH ARTICLE

WILEY

Hippocampal grading provides higher classification accuracy for those in the AD trajectory than hippocampal volume

Cassandra Morrison[1] | Mahsa Dadar[2,3] | Neda Shafiee[1] | D. Louis Collins[1] | for Alzheimer's Disease Neuroimaging Initiative

[1]McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

[2]Department of Psychiatry, McGill University, Montreal, Quebec, Canada

[3]Douglas Mental Health University Institute, Montreal, Quebec, Canada

Correspondence
Cassandra Morrison, Montreal Neurological Institute, 3801 University Street, Montreal, QC H3A 2B4, Canada.
Email: cassandra.morrison@mail.mcgill.ca

Funding information
Canadian Institutes of Health Research, Grant/Award Numbers: MFE-176608, PJT-165921; Alzheimer's Disease Neuroimaging Initiative; National Institutes of Health, Grant/Award Number: U01 AG024904; DOD ADNI, Grant/Award Number: W81XWH-12-2-0012; National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering

Abstract

Much research has focused on neurodegeneration in aging and Alzheimer's disease (AD). We developed Scoring by Nonlocal Image Patch Estimator (SNIPE), a non-local patch-based measure of anatomical similarity and hippocampal segmentation to measure hippocampal change. While SNIPE shows enhanced predictive power over hippocampal volume, it is unknown whether SNIPE is more strongly associated with group differences between normal controls (NC), early MCI (eMCI), late (lMCI), and AD than hippocampal volume. Alzheimer's Disease Neuroimaging Initiative older adults were included in the first analyses ($N = 1666$, 513 NCs, 269 eMCI, 556 lMCI, and 328 AD). Sub-analyses investigated amyloid positive individuals ($N = 834$; 179 NC, 148 eMCI, 298 lMCI, and 209 AD) to determine accuracy in those on the AD trajectory. We compared SNIPE grading, SNIPE volume, and Freesurfer volume as features in seven different machine learning techniques classifying participants into their correct cohort using 10-fold cross-validation. The best model was then validated in the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL). SNIPE grading provided the highest classification accuracy for all classifications in both the full and amyloid positive sample. When classifying NC:AD, SNIPE grading provided an 89% accuracy (full sample) and 87% (amyloid positive sample). Freesurfer volume provided much lower accuracies of 65% (full sample) and 46% (amyloid positive sample). In the AIBL validation cohort, SNIPE grading provided a 90% classification accuracy for NC:AD. These findings suggest SNIPE grading provides increased classification accuracy over both SNIPE and Freesurfer volume. SNIPE grading offers promise to accurately identify people with and without AD.

KEYWORDS
Alzheimer's disease, classification accuracy, Freesurfer, hippocampal grading, hippocampal volume, mild cognitive impairment, older adults

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Data was also used from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was also available from the ADNI database. As such, the investigators within ADNI/AIBL contributed to the design and implementation of ADNI/AIBL protocols and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. A complete list of AIBL researchers is listed at www.aibl.csiro.au.

1 | INTRODUCTION

Dementia is a term used to describe a range of disorders that are caused by abnormal brain changes in aging. These abnormal changes impair cognitive functions such as memory, language, and problem-solving that are severe enough to interfere with daily life and independence (Alzheimer's Association, 2021). The most common form of dementia is Alzheimer's disease (AD), a progressive neurodegenerative disorder defined by its underlying pathologies of β amyloid (Aβ42) deposition, pathological tau, and neurodegeneration [AT(N)] (Jack, 2018). Unfortunately, these pathological changes can start years, even decades, before the onset of cognitive symptoms (Sperling et al., 2011). Researchers must thus develop diagnostic tools that can detect AD pathology before too much irreversible neurodegeneration occurs.

Many recent studies have attempted to improve AD classification accuracy using various biomarkers such as cognitive testing, positron emission tomography (PET), cerebrospinal fluid (CSF) assays of Aβ42, tau, or magnetic resonance imaging (MRI) changes. For example, using episodic memory test such as the California Verbal Learning Test (CVLT) or Rey Auditory Verbal Learning Test (RAVLT) yield accuracies of over 80% when predicting conversion from mild cognitive impairment (MCI) to AD (Eckerström et al., 2013; Rabin et al., 2009). However, predicting future progression or diagnosis from NC (or even MCI and dementia) is difficult due to clinician subjectivity and individual patient variability. The implementation of machine learning techniques to analyze AD-related biomarkers may help improve early detection models and increase classification accuracy. Furthermore, the use of imaging techniques as opposed to cognitive tests may improve accuracies because they are not influenced by clinician subjectivity and less influenced by patient day-to-day variability.

Using MRI, researchers can measure neurodegeneration by analyzing volumetric changes in older adults' brains. When studying changes due to aging, MCI, and AD, the most commonly studied area is the hippocampus because this region experiences atrophy early in the disease course (Fjell et al., 2014). The hippocampus has also been identified as one of the most useful biomarkers when examining progression from MCI to AD (Risacher et al., 2009). Another method of measuring hippocampal differences between groups is hippocampal grading, measured by the Scoring by Nonlocal Image Patch Estimator (SNIPE) (Coupé et al., 2012; Coupé et al., 2012; Coupé et al., 2015).

SNIPE computes the similarity of every voxel in the hippocampus of each person to a large library of manually segmented MRI datasets from both healthy cognitively intact older adults and an equal number of patients with AD. This procedure compares the local neighborhood patch surrounding the voxel to corresponding neighborhood patches for each image volume in the library. The SNIPE score is the average of similarity-weighted labels (i.e., −1 for AD and +1 for healthy control) from the library of individuals. When the average SNIPE score is positive, the structure is more similar to healthy control, and when negative, the structure is more similar to AD. The SNIPE hippocampal segmentation method is based on the Pruessner et al. (2000) anatomical protocol, which is very similar to the European Alzheimer's Disease Consortium (EADC) harmonized protocol designed to maximize the difference in hippocampal volume for older adults with AD (Frisoni et al., 2015). SNIPE grading is a similarity function that not only takes into account volume, but also computes the similarity of the entire structure including the texture, intensity, and shape in its estimate of whether an individuals' hippocampus is more similar to the AD or NC template (Coupé et al., 2012). Previous research techniques have found that in addition to volume, both shape (Shen et al., 2012) and texture (Sørensen et al., 2017) are important factors to consider when examining memory decline and AD. The benefit of SNIPE grading over SNIPE volume is that grading integrates several important features (i.e., volume, shape, and texture) into one measure to provide valuable information that cannot be obtained from volume alone.

The additional features included in the SNIPE grading metric contribute to the strong associations between grading and cognition and AD. For example, grading has been shown to be associated with episodic memory changes in people with subjective cognitive decline and global cognition changes in cognitively healthy older adults and people with early MCI (Morrison et al., 2023). Furthermore, this method has proven to classify between cognitively healthy older adults and people with AD with an accuracy of 93% when using both the hippocampus and entorhinal cortex (Coupé et al., 2012). However, these SNIPE results come from limited samples and need to be further examined to determine their usefulness in classifying different disease cohorts.

While automated classification of AD and MCI is widely studied there are several limitations in the current research. To better understand the potential of SNIPE to correctly classify aging and cognitive impairment groups more research is needed on (1) larger samples and in people with MCI, (2) people who are amyloid positive and are thus on the AD trajectory, (3) comparing SNIPE grading and volume to established methods such as Freesurfer, and (4) validating the results in an external cohort to determine technique generalizability. A larger sample is needed in these machine learning studies to reduce the chance of overfitting and to improve generalizability to other samples. Coupé and colleagues' papers examining SNIPE used the Alzheimer's Disease Neuroimaging Initiative diagnosis label "AD" to differentiate groups (2012, 2015). The problem with the ADNI AD diagnosis is that this diagnosis is based solely on clinical scores and does not include amyloid positivity. Based on the National Institute on Aging-Alzheimer's Association biomarker AD profiles, an older adult with abnormal amyloid levels (amyloid positive) is placed in the Alzheimer's continuum, whereas someone without amyloid positivity is either experiencing non-AD pathologic change or has normal-level AD biomarkers (Jack, 2018). Thus, in order to correctly classify people as AD or as NC on the disease continuum it is important to examine only those who are amyloid positive. Specifically examining people on the AD continuum is also necessary because they are often selected for clinical trials and their accurate classification has the potential to further improve clinical trial patient selection. Finally, while the current research on SNIPE grading and volume has offered promising results

(Coupé et al., 2012; Coupé et al., 2012; Coupé et al., 2015), SNIPE grading methods have yet to be compared to the traditional methods used to measure hippocampal volume (i.e., Freesurfer).

In order for the SNIPE method to be applicable in clinical research, clinical trials, or clinical settings it is necessary to determine whether SNIPE grading or volume provide higher classification accuracy than current standards in the field. Therefore, the goal of this study was to compare classification accuracy of SNIPE hippocampal grading and SNIPE hippocampal volume measures to Freesurfer volume. A recent review has also shown that few studies compute classifications between MCI vs. AD, with most studies focusing on classifying NC from AD (Tanveer et al., 2020). While the former is a bit late for early intervention, the latter is not really of clinical interest. We thus designed this study to examine classification accuracy between healthy older adults, people early mild cognitive impairment (eMCI) and late MCI (lMCI), and people with AD. Importantly, we also completed these analyses separately in amyloid positive individuals, focusing on people who are in the earliest stages of pathological AD, are most likely to convert to AD, and are most likely to be selected for clinical trials. Several commonly used classifiers were applied to determine which technique (i.e., SNIPE Grading, SNIPE volume or FreeSurfer volume) is best at classifying participants into their correct diagnostic cohort. Including multiple techniques improves the generalizability of our results and comparison to past research.

2 | METHODS

2.1 | Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Participants were selected from ADNI-1, ADNI-2, and the ADNI-GO cohorts. The study received ethical approval from the review boards of all participating institutions. Written informed consent was obtained from participants or their study partner.

All ADNI participants were imaged using a 3T scanner with T1-weighted imaging parameters (see http://adni.loni.usc.edu/methods/mri-tool/mri-analysis/ for the detailed MRI acquisition protocol). Baseline scans were downloaded from the ADNI public website.

2.2 | Participants from ADNI cohort

Participant inclusion and exclusion criteria are available at www.adni-info.org. All participants were between the ages of 55 and 90 at the time of recruitment, exhibiting no evidence of depression. Healthy normal controls had no evidence of memory decline, as measured by the Wechsler Memory Scale and no evidence of impaired global cognition as measured by the Mini Mental Status Examination (MMSE) or Clinical Dementia Rating (CDR). Both eMCI and lMCI had scores between 24 and 30 on the MMSE, 0.5 on the CDR, and abnormal scores on the Wechsler Memory Scale. AD was defined as participants who had abnormal memory function on the Wechsler Memory Scale, an MMSE score between 20 and 26, a CDR of 0.5 or 1.0 and probable AD according to the NINCDS/ADRDA criteria.

Figure 1 summarizes the methodology used to select participants from the ADNI studies. A total of 1666 participants were selected from ADNI-1 ($n = 799$), ADNI-2 ($n = 776$), ADNI-GO ($n = 91$) who had baseline MRI scans that passed quality control, for which hippocampal grading and volume could be extracted (513 NC, 269 eMCI, 556 lMCI, 328 AD).

Amyloid status was derived from PET or CSF measures. ADNI PET data was acquired on multiple PET instruments with different acquisition sequences following various platform-specific acquisition protocols. All PET data underwent quality control and standard image pre-processing correct steps to improve data uniformity across collection sites. More detail can be downloaded from the ADNI procedures manual. The AV-45 PET scans were collected approximately 50 min post injection (Landau & Jagust, 2015). The PiB-PET scans were collected after 50–70 min post injection of approximately 15 mCi (Jagust et al., 2010). To obtain cerebrospinal fluid (CSF) samples, lumbar punctions were performed as described in the ADNI procedures manual. CSF Aβ42 were measured using the multiplex xMAP Luminex platform (Luminex Corp, Austin, TX, USA) with the INNO-BIA AlzBio3 kit (Innogenetics) (Olsson et al., 2005; Shaw et al., 2009).

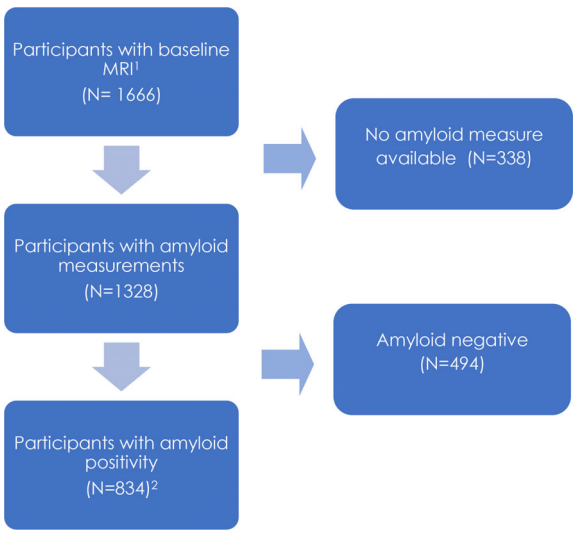

FIGURE 1 Flowchart summarizing the participant inclusion and exclusion criteria based on amyloid positivity. [1]The participants included in the first analysis including the entire sample. [2]Amyloid positive participants included in the second analysis focusing on those in the AD trajectory.

The current definition of dementia in ADNI is solely based on clinical assessments and does not include amyloid positivity. For this reason, we also wanted to repeat the same analysis in a subset of amyloid positive participants to ensure we are testing on people who are experiencing Alzheimer's related pathological changes. To determine amyloid positivity, both CSF and PET values were used as not all participants had both measurements available. Participants were identified as amyloid positive if they had any of the following at baseline: (1) a standardized update value ratio (SUVR) of >1.11 on AV45 PET (Landau et al., 2013), (2) a SUVR of >1.2 using Pittsburgh compound-B PET (Villeneuve et al., 2015), or (3) a cerebrospinal fluid Aβ1-42 ≤ 980 pg/mL as per ADNI recommendations. Of the 1666 participants selected, 1328 had baseline amyloid levels available to determine amyloid positivity (427 NC, 263 eMCI, 400 lMCI, 238 AD) and of those, 834 were amyloid positive (179 NC, 148 eMCI, 298 lMCI, 209 AD) and 494 were amyloid negative (248 NC, 115 eMCI, 102 lMCI, 29 dementia).

2.3 | Participants from AIBL cohort

Participant inclusion and exclusion criteria for the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) have been previously full described (Ellis et al., 2009). Briefly, healthy controls were 60+ and in good general health with no evidence of cognitive impairment. Those with MCI had to score < 28/30 on the MMSE, and have abnormal scores on the Wechsler Memory Scale, and a CDR score of 0.5 or greater. AD patients were characterized by the NINCDS-ADRDA criteria (McKhann et al., 1984). There were 858 participants, of these participants only 581 participants had baseline MRIs that passed quality control for which hippocampal grading and volume could be extracted were included (413 NC, 90 MCI, 78 AD).

2.4 | Structural MRI processing

Raw T1w scans for each participant were pre-processed through our standard pipeline including noise reduction (Coupé et al., 2008), intensity inhomogeneity correction (Sled et al., 1998), and intensity normalization into the range 0–100. The pre-processed images were then linearly (9 parameters: 3 translation, 3 rotation, and 3 scaling) (Dadar et al., 2018) registered to the MNI-ICBM152-2009c average (Fonov et al., 2011). The quality of the linear registrations was visually verified by an experienced rater (author M.D.), blinded to diagnostic group. Only seven datasets did not pass this quality control step and were discarded.

2.5 | SNIPE grading and volume

Scoring by Nonlocal Image Patch Estimator (SNIPE) was used to segment the hippocampus and measure the extent of AD-related change within the hippocampus using the linearly registered preprocessed T1-weighted images (Coupé et al., 2012; Coupé et al., 2012). The SNIPE procedure used has been previously described in detail (Dadar et al., 2020). In short, this technique uses a set of MRI volumes with manually segmented hippocampi as training library from both healthy aging subjects (CN) and patients with dementia due to AD. For each voxel from the subject under study that falls within a bounding box containing the medial temporal lobe region, a 3D $7 \times 7 \times 7$ patch centered around that voxel is compared with corresponding patches from the $N = 100$ MRI volumes (50 CN and 50 AD) in the training library. It is important to note that the 100 subjects used to create the training library were excluded from the analyses in the current paper. An intensity-based similarity metric (or "weight") between the patch under study and the training patch was then computed. These estimated weights were used to perform grading of the hippocampus based on the clinical label (CN vs. AD) of the training subjects:

$$g(x_i) = \frac{\sum_{s=1}^{N} \sum_{j \in \Omega} w(x_i, x_{s,j}).P_s}{\sum_{s=1}^{N} \sum_{j \in \Omega} w(x_i, x_{s,j})}$$

where x_i is the target voxel, and $g(x_i)$ is the corresponding grading value, and Ω is the search area. $w(x_i, x_{s,j})$ shows the similarity metric between surrounding patches of target voxel i and voxel j from training subject s. P_s is the clinical label of the training subject: we set it to -1 for AD patients and $+1$ for normal healthy subjects. This means that when a patch resembles CN anatomical characteristics more than AD, the grading score will be positive, conversely, if the patch is more similar to AD anatomy, the grading score is negative. The final SNIPE hippocampal grading score is an average of all the voxels within this structure in each hemisphere (see Figure 2 for an example pipeline of SNIPE).

In this method (SNIPE), volumes are calculated by counting voxels in a pseudo-Talairach stereotaxic space (ICBM152 template), thus correcting for subject's head size. The other benefit of SNIPE is that it addresses intersubject variability by using one-to-many mapping between each individual's anatomy and those of the training templates. The quality of the SNIPE segmentations were visually verified (author N.S.), blinded to diagnostic group. Only 11 datasets did not pass this quality control step and were discarded from the analyses.

2.6 | FreeSurfer volumes

Freesurfer volumes were used to complete the classification analysis with the ADNI data. The volumes were provided by ADNI, completed using the standard protocols developed and implemented by The University of California, San Francisco (UCSF). For a more detailed explanation of the pre-processing and quality control guidelines, please see the full UCSF FreeSurfer Overview and QC Guide.

2.7 | Classification analysis

To assess the prediction power of each measure, a classification scheme with 10-fold cross validation was used. Specifically, SNIPE grading, SNIPE volume, or Freesurfer volume scores for the left and right HC were summed and used as features along with participant age and sex.

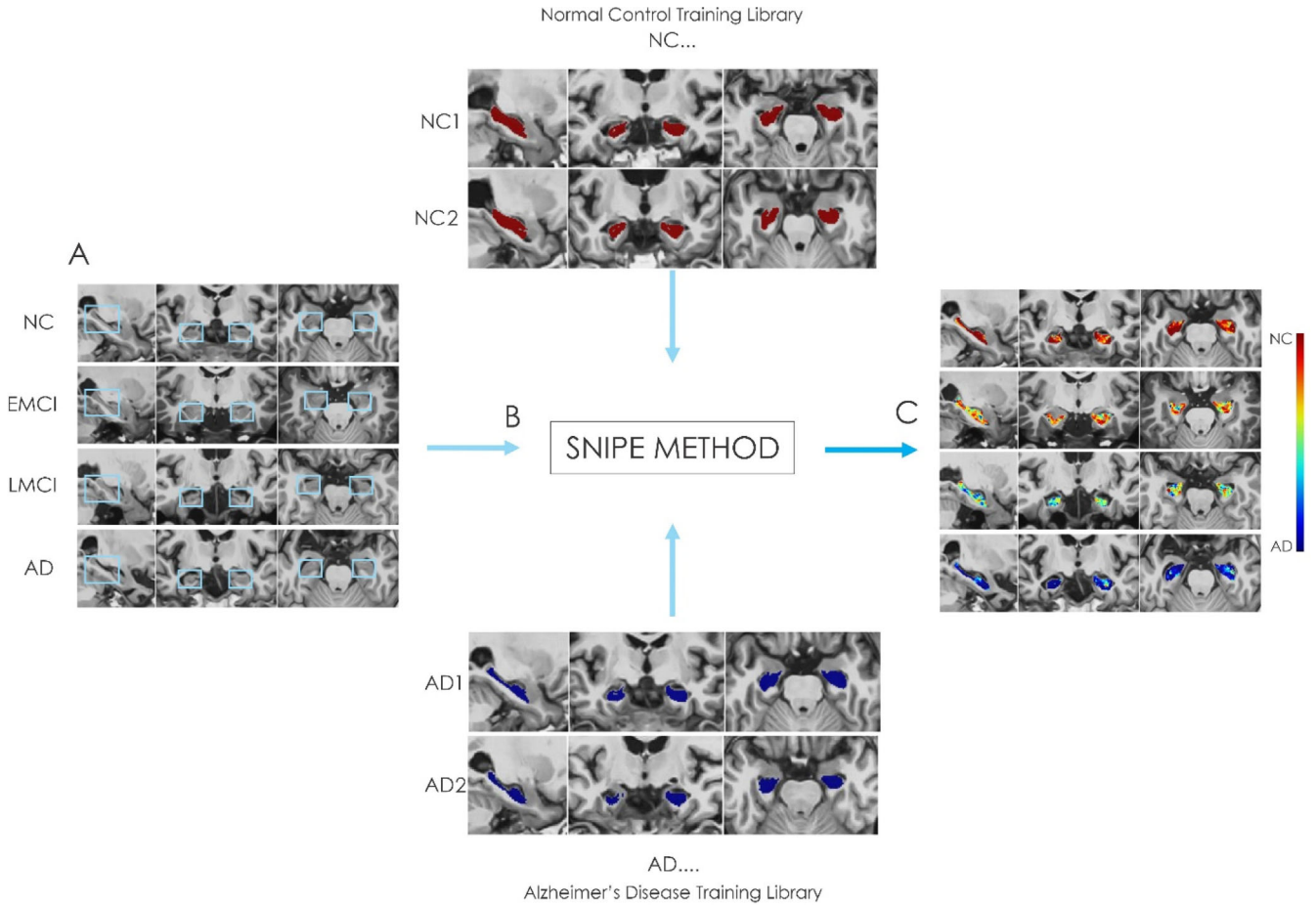

FIGURE 2 Overview of the SNIPE method. (a) New participant (either NC, eMCI, lMRI, or AD subject) MRI images input into the method. From a collection 50 NC subjects and 50 AD subjects in the training library, the *N* most similar NC templates and *N* most similar AD templates are selected for SNIPE processing (where $2N < 100$). A very generous medial temporal lobe mask (not shown) is used to limit the number of voxels under consideration. For each voxel within this mask, the blue outline represents an example 3D $7 \times 7 \times 7$ voxel patch that is centered around the specific voxel. (b) This patch is then compared and matched to nearby corresponding patches from pre-selected NC and AD subjects. (c) The voxel grading score is an average over matching patches from NC (red or $+1$) or AD (blue or -1), weighted by the patch similarity (rightmost images show grading maps for typical NC, eMCI, lMRI, or AD subjects). The grading scores are averaged over all voxels to yield a SNIPE grading for the left and the right hippocampus. NC, normal control. EMCI, early mild cognitive impairment. LMCI, late mild cognitive impairment. AD, Alzheimer's disease. NC1 to NC_N, normal control participants, preselected from the training library. AD1 to AD_N, Alzheimer's disease participants, preselected from the training library.

To ensure that the potential differences in the distribution of the random splits in the cross-validation folds do not impact the results, the same splits were consistently used for assessment of the performance of the three features evaluated with a support vector machine (SVM) classifier implemented within Scikit-Learn (https://scikit-learn.org/). The default parameters were used for this classifier including: Bayesian optimization (bayesop), Acquisition Function Name was "expected-improvement-per-second-plus," Empirical prior probability, 10 grid divisions were used (i.e., NumGridDivisions), and 30 was the maximum number of objective function evaluations (i.e., MaxObjectiveEvaluations). To facilitate comparison with other studies in the literature and to ensure that the results are not classifier dependent, six other classifiers were examined: decision tree, error-correcting output codes, binary Gaussian kernel, binary linear, and random forests. The default parameters were used for each classifier and are presented in Table S1. All analyses were performed using MATLAB version 9.7.

Independent validation of the classification was completed using NC, MCI, and AD participants from the AIBL cohort. While ADNI classifies MCI participants into either eMCI or lMCI, AIBL only uses MCI. For that reason classification models for NC:MCI and MCI:AD were created using only the ADNI training set: (1) with both eMCI and lMCI participants, (2) with only the eMCI participants, and (3) with only the lMCI participants. Importantly, AIBL data was used only for independent validation of these three models. No AIBL data was used in the creation of these models, nor for any parameter adjustment.

2.8 | Statistical analysis

Demographic information between groups was compared using independent sample *t*-tests and corrected for multiple comparisons using Bonferroni correction.

3 | RESULTS

3.1 | Demographic and clinical results in ADNI

Table 1 provides the demographic characteristics for all participants separated by group for both the full sample and amyloid positive subgroup.

Several demographic and clinical factors differed between subgroups in the full sample after correction for multiple comparison. With respect to age differences, eMCI were significantly younger than NC ($t = 6.84$, $p < .001$), people with lMCI ($t = 5.74$, $p < .001$), and people with AD ($t = 6.71$, $p < .001$). AD had significantly lower education than NC ($t = 5.84$, $p < .001$), people with eMCI ($t = 3.54$, $p < .001$), and lMCI ($t = 3.59$, $p < .001$). All groups significantly differed in ADAS-13 scores, with scores progressively increasing with each stage of decline (NC:eMCI, $t = -8.31$, $p < .001$; eMCI:lMCI, $t = -13.89$, $p < .001$; and lMCI:AD, $t = -21.51$, $p < .001$). Similarly, all groups significantly differed in CDR-SB scores, with scores progressively increasing with each stage of decline (NC:eMCI, $t = -26.44$, $p < .001$; eMCI:lMCI, $t = -6.01$, $p < .001$; and lMCI:AD, $t = -28.51$, $p < .001$).

Demographic differences were also observed in the amyloid positive sub-analysis. After correction for multiple comparisons, eMCI were younger than only NC ($t = 3.49$, $p < .001$). No group differences were observed in education after correction for multiple comparisons. All groups significantly differed in ADAS-13 scores, with scores progressively increasing with each stage of decline (NC:eMCI, $t = -7.12$, $p < .001$; eMCI:lMCI, $t = -10.16$, $p < .001$; and lMCI:AD, $t = -15.83$, $p < .001$). Similarly, all groups significantly differed in CDR-SB scores, with scores progressively increasing with each stage of decline (NC: eMCI, $t = -18.34$, $p < .001$; eMCI:lMCI, $t = -3.49$, $p < .001$; and lMCI:AD, $t = -24.07$, $p < .001$).

3.2 | Classification results in ADNI

Table 2 shows the participant classification accuracy, sensitivity, and specificity scores for SNIPE grading, SNIPE volume, and Freesurfer volume for each analysis using SVM. Accuracy, sensitivity, and specificity obtained using the other classifiers are available in Table S2.

The highest classification accuracy was obtained using SNIPE grading, followed by SNIPE volume, then Freesurfer. This order was obtained for all group classifications in both the full sample and amyloid positive sample. When examining NC vs. AD classification, SNIPE grading provided an accuracy of 89% (±4) compared to 80% (±4) for SNIPE volume and 65% (±12) obtained using Freesurfer. When examining NC vs. AD in the amyloid positive sample SNIPE grading accuracy dropped by only 2% to 87% (±2). On the other hand, SNIPE volume accuracy dropped by 4% to 76% (±5) and Freesurfer classification accuracy dropped to 46% (±11), a difference of 19%. In the NC vs. eMCI classification, SNIPE grading obtained the highest with 70% (±4) accuracy compared to SNIPE volume with 68% (±3) and Freesurfer volume with only 62% (±6). When comparing NC vs. eMCI in the amyloid positive sample SNIPE grading obtained the highest accuracy compared to SNIPE volume and Freesurfer volume (63% ± 6 vs. 59% ± 5 and 56% ± 10, respectively).

In the eMCI vs. lMCI classification, SNIPE grading and SNIPE volume obtained the highest with 67% (±5) accuracy compared to

TABLE 1 Demographic information for cognitively normal, early and late MCI, and AD participants.

	NC	eMCI	lMCI	AD
Full sample	***n* = 513**	***n* = 269**	***n* = 556**	***n* = 328**
Age	74.36 ± 5.79	70.82 ± 7.38	73.99 ± 7.53	75.00 ± 7.78
Education	16.35 ± 2.71	15.93 ± 2.65	15.90 ± 2.92	15.16 ± 3.01
Female sex	265 (52%)	120 (45%)	215 (39%)	147 (45%)
ADAS-13	9.30 ± 4.54	12.59 ± 5.54	18.70 ± 6.53	29.95 ± 7.90
CDR-SB	0.05 ± 0.19	1.29 ± 0.75	1.65 ± 0.92	4.45 ± 1.62
	NC	**eMCI**	**lMCI**	**AD**
Amyloid positive	***n* = 179**	***n* = 148**	***n* = 298**	***n* = 209**
Age	74.81 ± 5.68	72.29 ± 7.11[a]	73.51 ± 7.14	74.16 ± 8.01
Education	16.22 ± 2.67	15.83 ± 2.74	16.00 ± 2.87	15.44 ± 2.79
Female sex	86 (62%)	62 (42%)	121 (41%)	93 (44%)
ADAS-13	9.64 ± 4.62	13.70 ± 5.46	19.77 ± 6.65	30.74 ± 8.21
CDR-SB	0.06 ± 0.27	1.40 ± 0.84	1.70 ± 0.91	4.53 ± 1.60

Note: Values are expressed as mean ± standard deviant, or number (percentage %). Female sex is represented as total number of sample and percentage of sample.
Abbreviations: AD, Alzheimer's disease; ADAS-13, Alzheimer's Disease Assessment Scale–Cognitive Subscale; CDRSB, Clinical Dementia Rating Scale—Sum of Boxes; eMCI, early mild cognitive impairment; lMCI, late mild cognitive impairment; NC, cognitively normal controls.
[a]eMCI were younger than NC.

TABLE 2 Percent accuracy, sensitivity, and specificity (± standard deviation, evaluated over 10 folds) for classifying cognitively normal, early and late MCI, and AD participants.

	NC:AD			NC:eMCI			eMCI:lMCI			lMCI:AD		
SVM	**Grading**	**Volume**	**Freesurfer**	**Grading**	**Volume**	**Freesurfer**	**Grading**	**Volume**	**Freesurfer**	**Grading**	**Volume**	**Freesurfer**
Accuracy	89 ± 4	80 ± 4	65 ± 12	70 ± 4	68 ± 3	62 ± 6	67 ± 5	67 ± 5	53 ± 14	68 ± 6	63 ± 5	60 ± 7
Sensitivity	82 ± 6	66 ± 6	30 ± 3	23 ± 8	15 ± 4	37 ± 2	89 ± 8	100 ± 0	49 ± 3	45 ± 6	0 ± 0	24 ± 3
Specificity	94 ± 5	89 ± 8	87 ± 8	94 ± 4	96 ± 2	76 ± 17	23 ± 2	0 ± 0	60 ± 2	82 ± 5	100 ± 0	82 ± 2
Amyloid positive	**Grading**	**Volume**	**Freesurfer**	**Grading**	**Volume**	**Freesurfer**	**Grading**	**Volume**	**Freesurfer**	**Grading**	**Volume**	**Freesurfer**
Accuracy	87 ± 2	76 ± 5	46 ± 11	63 ± 6	59 ± 5	56 ± 10	67 ± 5	67 ± 4	51 ± 14	67 ± 4	58 ± 4	57 ± 4
Sensitivity	85 ± 5	77 ± 5	26 ± 3	52 ± 6	48 ± 2	32 ± 2	88 ± 8	100 ± 0	44 ± 3	57 ± 9	0 ± 0	23 ± 3
Specificity	89 ± 3	75 ± 5	68 ± 9	72 ± 4	69 ± 1	77 ± 2	27 ± 2	0 ± 0	65 ± 3	74 ± 6	100 ± 0	81 ± 15

Note: Grading = Scoring by Nonlocal Image Patch Estimator Hippocampal Grading, Volume = the Scoring by Nonlocal Image Patch Estimator Hippocampal Volume, Freesurfer = Hippocampal volume measured with Freesurfer.

Abbreviations: AD, Alzheimer's disease; eMCI, early mild cognitive impairment; lMCI, late mild cognitive impairment; NC, normal controls; SVM, support vector machines.

Freesurfer volume with 53% (±14). When comparing eMCI vs. lMCI in the amyloid positive sample SNIPE grading and volume accuracies did not change while Freesurfer accuracy dropped 2 to 51% (±14). In the lMCI vs. AD classification, SNIPE grading obtained the highest with 68% (±6) accuracy compared to SNIPE volume at 63% (±5) and Freesurfer with 60% (±7). When comparing lMCI vs. AD in the amyloid positive sample SNIPE grading once again provided the highest classification accuracy compared to SNIPE volume and Freesurfer volume (67% ± 4 vs. 58% ± 4 and 57% ± 4, respectively).

It is interesting to note that in all experiments, the standard deviation of the Freesurfer volume classification accuracy is much larger that SNIPE grading or SNIPE volumes except for lMCI:AD.

3.3 | Classification results in the AIBL cohort (external validation)

In the first set of classification results, we observed that SNIPE grading and volume were both more accurate than using hippocampal volumes obtained using Freesurfer at classifying groups. For that reason, in this validation procedure only SNIPE grading and volume were compared because both these techniques outperformed hippocampal volumes obtained using Freesurfer. Furthermore, given the similar prediction accuracies between the different classifiers, this external validation was completed using only the SVM classifier. Table 3 shows the participant classification accuracy, sensitivity, and specificity scores for SNIPE grading and SNIPE volume for each analysis using SVM.

The highest classification accuracy was obtained using SNIPE grading over SNIPE volume for almost all analyses. When examining NC vs. AD classification, SNIPE grading provided an accuracy of 90% compared to 80% for SNIPE volume. In the NC:MCI prediction, accuracy for SNIPE grading was higher than SNIPE volume 55% vs. 41% (trained with eMCI and lMCI) and 74% vs. 64% (trained with lMCI only), but higher for SNIPE volume over SNIPE grading 82% vs. 79% (trained with eMCI only). In the MCI:AD prediction, accuracy for SNIPE grading was higher than SNIPE volume in all three training cases, 66% vs. 54% (trained with eMCI and lMCI), 65% vs. 59% (trained with eMCI only), 66 vs. 54% (trained with lMCI only).

3.4 | Correlation analysis

To further explore why SNIPE grading and volume exhibited improved accuracy over Freesurfer volume, additional exploratory correlations were completed to examine the strength of the association between both SNIPE measures (volume and grading) to Freesurfer volume individually for each group. These findings are presented in Table 4. All significant associations between SNIPE and Freesurfer showed only a moderate correlation. The only moderate correlation indicates that the different procedures (e.g., SNIPE vs. Freesurfer) are not measuring the same aspects of the hippocampus. This only moderate correlation may explain why the differences in classification accuracy were so large in some cases.

TABLE 3 Percent accuracy, sensitivity, and specificity for classifying NC, MCI and AD participants in the AIBL cohort.

SVM	NC:AD		NC:MCI[a]		NC:MCI[b]		NC:MCI[c]		MCI:AD[a]		MCI:AD[b]		MCI:AD[c]	
	Grading	Volume	Grading	Volume	Grading	Volume	Grading	Volume	Grading	Volume	Grading	Volume	Grading	Volume
Accuracy	90	80	55	41	79	82	74	64	66	54	65	59	66	54
Sensitivity	77	59	67	79	27	0	56	59	29	0	76	76	40	0
Specificity	93	88	53	32	90	100	78	65	93	100	57	46	89	100

Note: Grading = Scoring by Nonlocal Image Patch Estimator Hippocampal Grading, Volume = the Scoring by Nonlocal Image Patch Estimator Hippocampal Volume.

Abbreviations: AD, Alzheimer's disease; AIBL, Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing; MCI, mild cognitive impairment; NC, normal controls; SVM, support vector machines.

[a]Training sample from ADNI used both eMCI and lMCI.

[b]Training sample from ADNI used only eMCI.

[c]Training sample from ADNI used only lMCI.

4 | DISCUSSION

In recent years, numerous studies have implemented machine learning techniques on imaging data with the goal of accurately classifying and predicting people with dementia and more specifically AD (see Tanveer et al., 2020 for review). The current study expands on the current findings by comparing classification accuracy on a large sample of NC, and people with eMCI, lMCI, and AD (and with MCI and AD in AIBL), examining only those on the AD-trajectory by focusing on those who are amyloid positive, and by employing a relatively new method to detect hippocampal changes that incorporates texture, intensity, shape, and volume into one metric. The findings observed here show that SNIPE grading has higher classification accuracy than both SNIPE volume and Freesurfer volume when classifying: (1) NC:AD, (2) NC:eMCI, (3) eMCI:lMCI, and (4) lMCI:AD in both the full sample and the amyloid positive sub-sample of ADNI and when classifying (1) NC:AD, (2) NC:MCI, and (3) MCI:AD in the AIBL cohort.

These findings complement those previously completed on SNIPE grading and volume (Coupé et al., 2012; Coupé et al., 2012; Coupé et al., 2015). They found that SNIPE grading could classify between NC and AD with 93% accuracy using both hippocampal and entorhinal cortex grading (Coupé et al., 2012). Coupé et al. (2012) also observed that grading was more accurate than volume at classifying progressive MCI vs. stable MCI. The current study also found high accuracy (89%) when classifying NC vs. AD using only hippocampal grading and observed the novel finding that when focusing on those in the AD trajectory (amyloid positive) the results also remained high (87%). This finding of high accuracy in the amyloid positive group is important when attempting to differentiate those who are at different points on the AD trajectory. High accuracy at identifying those at different stages of the AD trajectory is important for clinical trials and further shows the importance of these results. Future work should determine the predictability of SNIPE at determining which amyloid positive NC will convert to pathological AD. It should be noted, however, that while we are examining people on the AD trajectory based on amyloid positivity, that we do make claims for the ability of SNIPE to differentiate between amyloid positive and amyloid negative individuals.

Compared to Freesurfer, SNIPE measures provided higher classification accuracy for all group classifications. The comparison of SNIPE to Freesurfer is essential in determining the usefulness of SNIPE because Freesurfer is a common method used to examine volume changes in AD. A quick MEDLINE search shows almost 600 papers with the keywords of Freesurfer and AD. Furthermore, Freesurfer volumetric measures are provided for the hippocampus in ADNI. While Freesurfer is somewhat accurate at classifying NC vs. AD (65%), grading was much more accurate (89%), with a 24% improvement in classification accuracy. Similarly, differentiating between NC vs. AD in amyloid positive group the classification was much more robust with grading compared to Freesurfer (87% vs. 52%), with grading offering 35% higher accuracy. These findings show that Grading is more sensitive than Freesurfer to hippocampal differences that occur in AD compared to NC. It should be noted that while we use the term AD to refer to both the full-sample based on clinical diagnosis and the

TABLE 4 Correlation between SNIPE grading, SNIPE volume, and Freesurfer Volume for the full ADNI sample.

	NC	eMCI	lMCI	AD
SNIPE grading vs. Freesurfer volume	$r = .52$, $t = 13.08$, $p < .001$	$r = .66$, $t = 13.73$, $p < .001$	$r = .73$, $t = 22.33$, $p < .001$	$r = .69$, $t = 15.40$, $p < .001$
SNIPE volume vs. Freesurfer volume	$r = .51$, $t = 12.65$, $p < .001$	$r = .63$, $t = 12.72$, $p < .001$	$r = .68$, $t = 19.67$, $p < .001$	$r = .67$, $t = 14.55$, $p < .001$

amyloid positive sub-sample, the full sample represents AD and other dementias. Thus, these findings suggest that not only is grading accurate at detecting to hippocampal changes due to dementia but is also highly accurate at classifying between NC and AD subjects who are on the AD path.

There are a few important benefits of using the SNIPE grading scores proposed here regarding the robustness of the method. In the 1666 MRI volumes processed, only 7 failed stereotaxic registration and 11 failed SNIPE segmentation. Robustness is important in clinical trials since losing data to pipeline failures results in reduced power to detect group differences. In addition to being more accurate than Freesurfer at detecting group differences, SNIPE also had a lower standard deviation. For example, when classifying NC vs. AD in the full sample Freesurfer had a standard deviation (SD) of 12% which is three times higher than the SNIPE grading SD at only 4%. This finding was also observed in the amyloid positive NC vs. AD classification (11% SD for Freesurfer and 2% SD for SNIPE). Higher SD was observed for Freesurfer in all classification analysis. Lower SD and higher accuracy using SNIPE is because this method takes into account macroscopic changes in anatomy reflected in MRI texture by computing similarity between a subject and a template library of NCs and people with AD. Essentially, this technique computes anatomic similarity of the medial temporal lobe structures between each subject and the NC and AD templates to determine if they are more similar to NC or AD. Using MRI texture is thus much more specific than volume measurements, which is why SNIPE grading is better than both Freesurfer volume and SNIPE volume. SNIPE volume was also observed to have higher accuracy than Freesurfer volume in almost all cases. This higher accuracy in SNIPE volume over Freesurfer volume may be related to the anatomical definition that drives the estimated volume. As mentioned, SNIPE is based off the Pruessner anatomical protocol which is similar to the EADC harmonized protocol which was specifically designed to maximize the difference observed between cognitively healthy older adults and people with Alzheimer's disease by including substructures known to have high NC:AD effect sizes. Therefore, the SNIPE method was designed to detect subtle differences between the groups that may not observed using Freesurfer.

Similar classification accuracies were also obtained when validating our results in the independent AIBL cohort. We obtained a high accuracy of 90% when classifying NC vs. AD in the AIBL cohort. Accuracies similar to those observed using the ADNI dataset were also observed in the AIBL NC vs. MCI and MCI vs. AD classification. These findings suggest that the classification model is not only accurate in the ADNI dataset but is generalizable to and replicable in other datasets. In a recent meta-analysis, authors found that almost 30% of articles reviewed use the test set in the training process, thus *double dipping* during their evaluation (Ansart et al., 2021). This finding further emphasizes the importance of the current study using cross-validation in the original dataset as well as using an independent cohort for validation.

The results found here are better or equivalent to past machine learning techniques that attempt to classify different disease cohorts from each other and NCs (see Tanveer et al., 2020). The majority of the studies examined in the above-mentioned review focused on only classifying NC vs. AD, with less than 25% of the 60 studies using SVM to classify between MCI and AD and just over 30% classifying between NC and MCI (Tanveer et al., 2020). In order to target early diagnosis of AD, researchers must be able to correctly classify between MCI and AD. In our sample, we were able to differentiate between eMCI vs. lMCI and lMCI vs. AD with almost 70% accuracy. Although there is some research that has shown similar success the novelty and advantage of the current results is that we employed a larger sample to improve generalizability, examined those on the AD trajectory (by studying amyloid positive groups), and classified people with eMCI vs. lMCI. Furthermore, we also validated these results in an independent cohort of NCs, MCI, and AD. These findings show promise for the use of SNIPE grading as a powerful MRI-based feature that could be used in conjunction with other data to improve classification of patients into the correct disease cohort. Future research should examine whether SNIPE grading is useful at detecting which amyloid positive subjects will cognitively decline and develop AD. Early detection of AD will not only improve a clinicians' ability to provide effective care to patients but also potentially improve selection of patients for clinical trials. It should be noted that a limitation of the current study was our inability to separate the AIBL participants based on amyloid positivity because this measure was not available within the LONI database at the time of this study. Therefore, our validation on the AIBL cohort does not consider amyloid status. This restriction is a limitation of this study because this analysis would not target those who are on the AD continuum.

5 | CONCLUSION

The current paper observed that SNIPE grading scores provided higher classification accuracy than both SNIPE volume and Freesurfer volumes. Importantly, this classification accuracy remained similar in the independent validation analysis using the AIBL cohort. These findings suggest that HC grading offers promise as a method to accurate classify those with and without AD. Future work should examine

whether HC grading is predictive of future conversion from NC to MCI and dementia.

ACKNOWLEDGMENTS

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

CONFLICT OF INTEREST STATEMENT

The authors report no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data used for this analysis (both AIBL and ADNI data) are available on request from the ADNI database (ida.loni.usc.edu).

FINANCIAL DISCLOSURES

Dr. Morrison is supported by a postdoctoral fellowship from Canadian Institutes of Health Research, Funding Reference Number: MFE-176608. Dr. Dadar reports receiving research funding from the Healthy Brains for Healthy Lives (HBHL), Alzheimer Society Research Program (ASRP), and Douglas Research Centre (DRC). Neda Shafiee has no conflicts of interest to declare. Dr. Collins reports receiving research funding from Canadian Institutes of Health research, the Canadian National Science and Engineering Research Council, Brain Canada, the Weston Foundation, and the Famille Louise & André Charron.

ORCID

Cassandra Morrison https://orcid.org/0000-0002-1087-5770

REFERENCES

Alzheimer's Association. (2021). *What is dementia*. https://www.alz.org/alzheimers-dementia/what-is-dementia

Ansart, M., Epelbaum, S., Bassignana, G., Bône, A., Bottani, S., Cattai, T., Couronné, R., Faouzi, J., Koval, I., Louis, M., Thibeau-Sutre, E., Wen, J., Wild, A., Burgos, N., Dormont, D., Colliot, O., & Durrleman, S. (2021). Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review. *Medical Image Analysis*, *67*, 101848.

Coupé, P., Eskildsen, S. F., Manjón, J. V., Fonov, V. S., Collins, D. L., & Alzheimer's Disease Neuroimaging Initiative. (2012). Simultaneous segmentation and grading of anatomical structures for patient's classification: Application to Alzheimer's disease. *NeuroImage*, *59*(4), 3736–3747. https://doi.org/10.1016/j.neuroimage.2011.10.080

Coupé, P., Eskildsen, S. F., Manjón, J. V., Fonov, V. S., Pruessner, J. C., Allard, M., Collins, D. L., & Alzheimer's Disease Neuroimaging Initiative. (2012). Scoring by nonlocal image patch estimator for early detection of Alzheimer's disease. *NeuroImage: Clinical*, *1*(1), 141–152. https://doi.org/10.1016/j.nicl.2012.10.002

Coupé, P., Fonov, V. S., Bernard, C., Zandifar, A., Eskildsen, S. F., Helmer, C., Manjón, J. V., Amieva, H., Dartigues, J.-F., Allard, M., Catheline, G., Collins, D. L., & Alzheimer's Disease Neuroimaging Initiative. (2015). Detection of Alzheimer's disease signature in MR images seven years before conversion to dementia: Toward an early individual prognosis. *Human Brain Mapping*, *36*(12), 4758–4770. https://doi.org/10.1002/hbm.22926

Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., & Barillot, C. (2008). An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE Transactions on Medical Imaging*, *27*(4), 425–441.

Dadar, M., Fonov, V. S., Collins, D. L., & Alzheimer's Disease Neuroimaging Initiative. (2018). A comparison of publicly available linear MRI stereotaxic registration techniques. *NeuroImage*, *174*, 191–200. https://doi.org/10.1016/j.neuroimage.2018.03.025

Dadar, M., Gee, M., Shuaib, A., Duchesne, S., & Camicioli, R. (2020). Cognitive and motor correlates of grey and white matter pathology in Parkinson's disease. *NeuroImage: Clinical*, *27*, 102353. https://doi.org/10.1016/j.nicl.2020.102353

Eckerström, C., Olsson, E., Bjerke, M., Malmgren, H., Edman, Å., Wallin, A., & Nordlund, A. (2013). A combination of neuropsychological, neuroimaging, and cerebrospinal fluid markers predicts conversion from mild cognitive impairment to dementia. *Journal of Alzheimer's Disease*, *36*(3), 421–431. https://doi.org/10.3233/JAD-122440

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., & AIBL Research Group. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International psychogeriatrics*, *21*(4), 672–687.

Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M., Walhovd, K. B., & Alzheimer's Disease Neuroimaging Initiative. (2014). What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Progress in Neurobiology*, *117*, 20–40. https://doi.org/10.1016/j.pneurobio.2014.02.004

Fonov, V., Evans, A., Botteron, K., Almli, C., McKinstry, R., & Collins, D. B. D. C. G. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327. https://doi.org/10.1016/j.neuroimage.2010.07.033

Frisoni, G. B., Jack, C. R., Jr., Bocchetta, M., Bauer, C., Frederiksen, K. S., Liu, Y., Preboske, G., Swihart, T., Blair, M., Cavedo, E., Grothe, M. J., Lanfredi, M., Martinez, O., Nishikawa, M., Portegies, M., Stoub, T., Ward, C., Apostolova, L. G., Ganzola, R., … Winblad, B. (2015). The EADC-ADNI harmonized protocol for manual hippocampal segmentation on magnetic resonance: Evidence of validity. *Alzheimer's & Dementia*, *11*(2), 111–125. https://doi.org/10.1016/j.jalz.2014.05.1756

Jack, C. R. (2018). NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Physiology & Behavior*, *176*(5), 139–148. https://doi.org/10.1016/j.jalz.2018.02.018.NIA-AA

Jagust, W. J., Bandy, D., Chen, K., Foster, N. L., Landau, S. M., Mathis, C. A., Price, J. C., Reiman, E. M., Skovronsky, D., Koeppe, R. A., & Alzheimer's Disease Neuroimaging Initiative. (2010). The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer's & Dementia*, *6*(3), 221–229. https://doi.org/10.1016/j.jalz.2010.03.00

Landau, S., & Jagust, W. (2015). *Florbetapir processing methods*. https://adni.bitbucket.io/reference/docs/UCBERKELEYAV45/ADNI_AV45_Methods_JagustLab_06.25.15.pdf. Helen Wills Neuroscience Institute, UC Berkeley and Lawrence Berkeley National Laboratory California.

Landau, S. M., Lu, M., Joshi, A. D., Pontecorvo, M., Mintun, M. A., Trojanowski, J. Q., Shaw, L. M., Jagust, W. J., & Alzheimer's Disease Neuroimaging Initiative. (2013). Comparing positron emission tomography imaging and cerebrospinal fluid measurements of β-amyloid. *Annals of Neurology*, *74*(6), 826–836. https://doi.org/10.1002/ana.23908

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939–939.

Morrison, C., Dadar, M., Shafiee, N., Collins, D. L., & Alzheimer's Disease Neuroimaging Initiative. (2023). The use of hippocampal grading as a biomarker for preclinical and prodromal Alzheimer's disease. *Human Brain Mapping*, *44*, 3147–3157. https://doi.org/10.1002/hbm.26269

Olsson, A., Vanderstichele, H., Andreasen, N., De Meyer, G., Anders, W., Holmberg, B., Rosengren, L., Vanmechelen, E., & Blennow, K. (2005). Simultaneous measurement of β-amyloid (1–42), total tau, and phosphorylated tau (Thr181) in cerebrospinal fluid by the xMAP technology. *Clinical Chemistry*, *51*(2), 336–345. https://doi.org/10.1373/clinchem.2004.039347

Pruessner, J. C., Li, L. M., Serles, W., Pruessner, M., Collins, D. L., Kabani, N., Lupien, S., & Evans, A. C. (2000). Volumetry of hippocampus and amygdala with high-resolution MRI and three-dimensional analysis software: Minimizing the discrepancies between laboratories. *Cerebral Cortex*, *10*(4), 433–442. https://doi.org/10.1093/cercor/10.4.433

Rabin, L. A., Paré, N., Saykin, A. J., Brown, M. J., Wishart, H. A., Flashman, L. A., & Santulli, R. B. (2009). Differential memory test sensitivity for diagnosing amnestic mild cognitive impairment and predicting conversion to Alzheimer's disease. *Aging, Neuropsychology, and Cognition*, *16*(3), 357–376. https://doi.org/10.1080/13825580902825220

Risacher, S. L., Saykin, A. J., Wes, J. D., Shen, L., Firpi, H. A., & McDonald, B. C. (2009). Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Current Alzheimer Research*, *6*(4), 347–361. https://doi.org/10.2174/156720509788929273

Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V. M.-Y., Trojanowski, J. Q., & Alzheimer's Disease Neuroimaging Initiative. (2009). Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of Neurology*, *65*(4), 403–413. https://doi.org/10.1002/ana.21610

Shen, K. K., Fripp, J., Mériaudeau, F., Chételat, G., Salvado, O., Bourgeat, P., & Alzheimer's Disease Neuroimaging Initiative. (2012). Detecting global and local hippocampal shape changes in Alzheimer's disease using statistical shape models. *NeuroImage*, *59*(3), 2155–2166. https://doi.org/10.1016/j.neuroimage.2011.10.014

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, *17*(1), 87–97. https://doi.org/10.1109/42.668698

Sørensen, L., Igel, C., Pai, A., Balas, I., Anker, C., Lillholm, M., Nielsen, M., & Alzheimer's Disease Neuroimaging Initiative. (2017). Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage: Clinical*, *13*, 470–482. https://doi.org/10.1016/j.nicl.2016.11.025

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Jr., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., … Phelps, C. H. (2011). 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. *Alzheimer's & Dementia*, *7*(3), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003

Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(1s), 1–35. https://doi.org/10.1145/3344998

Villeneuve, S., Rabinovici, G. D., Cohn-Sheehy, B. I., Madison, C., Ayakta, N., Ghosh, P. M., la Joie, R., Arthur-Bentil, S. K., Vogel, J. W., Marks, S. M., Lehmann, M., Rosen, H. J., Reed, B., Olichney, J., Boxer, A. L., Miller, B. L., Borys, E., Jin, L. W., Huang, E. J., … Jagust, W. (2015). Existing Pittsburgh compound-B positron emission tomography thresholds are too high: Statistical and pathological evaluation. *Brain*, *138*(7), 2020–2033. https://doi.org/10.1093/brain/awv11

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Morrison, C., Dadar, M., Shafiee, N., Collins, D. L., & for Alzheimer's Disease Neuroimaging Initiative (2023). Hippocampal grading provides higher classification accuracy for those in the AD trajectory than hippocampal volume. *Human Brain Mapping*, *44*(12), 4623–4633. https://doi.org/10.1002/hbm.26407