Brain MR radiomics to differentiate cognitive disorders

Sara Ranjbar, PhD¹, Stefanie N. Velgos, MSc², Amylou C. Dueck, PhD³, Yonas E. Geda, MD, MSc⁴,⁵, J. Ross Mitchell, PhD⁶, #, and The Alzheimer’s Disease Neuroimaging Initiative¶

1. Department of Research, Mayo Clinic Arizona
2. Center for Clinical and Translational Science, Mayo Clinic Graduate School of Biomedical Sciences, Mayo Clinic Arizona
3. Department of Biostatistics, Mayo Clinic Arizona
4. Department of Psychiatry and Psychology, Mayo Clinic Arizona
5. Department of Neurology, Mayo Clinic Arizona
6. Department of Physiology and Biomedical Engineering, Mayo Clinic Arizona

# These authors contributed equally to this work.

Abstract

Subtle and gradual changes occur in the brain years prior to cognitive impairment due to age-related neurodegenerative disorders. We examined the utility of hippocampal texture analysis and volumetric features extracted from brain magnetic resonance (MR) data to differentiate between three cognitive groups (cognitively normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer’s disease (AD)), and neuropsychological Clinical Dementia Rating (CDR) scores. Data from 173 unique patients with 3T T1-weighted MR images from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database were analyzed. A variety of texture and volumetric features were extracted from bilateral hippocampal regions and were used to perform binary classification of cognitive groups and CDR scores. We used Diagonal Quadratic Discriminant Analysis (DQDA) in a leave one out cross validation scheme. Sensitivity, specificity, and area under the receiver operating characteristic curve (AUC) were used to assess the performance of models. Our results show promise for hippocampal texture analysis to distinguish between no impairment and early stages of impairment (AUCs of 0.86 for CN-MCI and 0.95 for CDR0-CDR1 models, respectively). Volumetric features were more successful at differentiating between no impairment and advanced stages of impairment (AUCs of 0.89 for CN-AD and 0.98 for CDR0-CDR2, respectively). MR radiomics may be a promising tool to classify various cognitive groups.
1. Introduction

The global dementia epidemic carries a widespread emotional and financial burden on patient families, caregivers, and society [1]. Currently, dementia of the Alzheimer’s type is the sixth-leading cause of death in the United States yet is the only disease among the top 10 causes of death that cannot be prevented or cured [2]. To date, clinical trials for Alzheimer’s disease (AD) therapeutics have been universally disappointing.

One significant factor for the slow progress is the lack of powerful, early detection methods of cognitive impairment. AD is characterized by the deposition of beta amyloid (Aβ) and hyperphosphorylated tau, resulting in plaques and neurofibrillary tangles, respectively. One hypothetical biomarker model describes the temporal order of disease stages as follows: 1) Aβ plaque accumulation; 2) neuronal injury; 3) brain structure atrophy; 4) memory loss; and, 5) general cognitive decline [3]. Clinical trials may fail because these neuropathological changes precede cognitive deficit manifestations by several decades [4–8]. Consequently, irreversible brain damage may have already occurred. Thus, identifying quantifiable biomarkers for early cognitive impairment is of profound public health importance. Early detection may allow earlier pharmacological interventions, when patients may be more responsive to treatments. In addition, early detection would allow patients to make conscious decisions about their situation (personal and property) if their underlying diseases lead to progression to dementia. However, as of now, early detection of cognitive impairment is challenging.

Multiple studies have used structural magnetic resonance (MR) imaging to predict Alzheimer’s disease [9–13]. Several studies found that local hippocampal and total brain volume are significantly reduced in AD and mild cognitive impairment (MCI) as compared to healthy elderly [14–23]. The hippocampus is affected early, and generally severely, in the AD pathological process [24]. Hippocampal volume is the most studied structural biomarker of AD and is used in the criteria for AD diagnosis [25]. In addition, prediction of MCI to AD conversion has been correlated with the rate and amount of hippocampal, medial temporal lobe, and total brain atrophy [26–31].

Biomedical texture analysis aims to quantitatively describe pixel/voxel intensity distributions and the interrelations of pixel intensities across multiple spatial scales. Texture analysis has been used previously in the context of AD [14, 28, 32–35]. Radiomics is an emerging approach to image analysis and refers to high-throughput extraction of quantitative features from radiological images in order to convert images into structured and mineable data [36–38]. Radiomics pipelines often employ a variety of texture analysis methods to provide a holistic representation of texture-based information of the image, or of regions of interest in the image. Radiomics-based models have previously revealed predictive and prognostic associations between images and clinical outcomes [36–38]. These models offer the
potential of capturing often overlooked or hidden information of underlying disease
dynamics. Our group has developed a radiomics texture analysis platform that has been
previously used to characterize gene expression patterns of brain cancer \{39, 40\}, to aid in
the diagnosis of head and neck cancers \{41, 42\} and breast cancer \{43\}.

The overall goal of this study is to differentiate between cognitive groups (CN, MCI, AD)
and clinical dementia rating (CDR) scores using MR-based texture and volume
measurements from the hippocampus. We hypothesize that changes in neuropsychological
function related to cognitive impairment have a radiological counterpart, detectable via
structural MRI. We also hypothesize that texture analysis will be sensitive enough to identify
early MRI structural hippocampal changes related to the early AD pathophysiological process,
which will be correlated with cognitive groups and CDR scores. Specifically, our objectives
are two-fold: to use MR radiomics features to 1) differentiate between cognitive groups
(cognitively normal (CN), Mild Cognitive Impairment (MCI), Alzheimer’s disease (AD));
and, 2) predict neuropsychological performance, quantified via Clinical Dementia Rating
(CDR) scores. The contributions of this study are: a) identification of MR-derived features
that could be used in detecting early cognitive impairment, b) assessing the use of a granular
measure of cognition assessment (such as CDR scores) compared to generic grouping for
predictive modeling, and c) comparing the utilities of volume and texture features in this
task.

2. Methods

2.1 ADNI Dataset

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 with the primary goal 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). We selected cases
from the shared image collection ADNI-1, a 5-year study with a cohort of 200 cognitively
normal (CN), 200 MCI and 400 AD cases \{44\}. The participants were divided into the
assigned CN, MCI, and AD groups and underwent 3T imaging at the following time points:
baseline, 6, 12, 18 (MCI only), and 24 months. We categorized participants into three
cognitive groups as assigned by ADNI-1: CN, MCI, and AD. Group specific inclusion
criteria are available on ADNI’s website under the General Procedures Manual or under
Study Design, Background & Rationale \{45, 46\}. Briefly, cognitively normal participants
have mini-mental state exam (MMSE) scores between 24–30 (inclusive), a CDR of 0, non-
depressed, non-MCI, and non-demented \{45\}. MCI participants have MMSE scores
between 24–30 (inclusive), a memory complaint, have objective memory loss measured by
education adjusted scores on Wechsler Memory Scale Logical Memory II, a CDR of 0.5,
absence of significant levels of impairment in other cognitive domains, essentially
persevered activities of daily living, and an absence of dementia \{45\}. AD participants have
MMSE scores between 20–26 (inclusive), CDR of 0.5–2, abnormal memory function
documented by scoring below the education adjusted cutoff on the Logical Memory II
subscale (Delayed Paragraph Recall) from the Wechsler Memory Scale, and meets the NINCDS/ADRDA criteria for probable AD\cite{45}.

2.2 Cognitive measures

The Clinical Dementia Rating (CDR) score is obtained through semi-structured interviews with patients and informants to evaluate six domains: memory, orientation, judgment and problem solving, community affairs, home & hobbies, and personal care \cite{47}. Patients are then classified on the following ordinal scales: 0 (no impairment), 0.5 (questionable impairment), 1.0 (mild dementia), 2.0 (moderate dementia), or 3.0 (severe dementia). Typically, a score of 0.5 is given to individuals with a diagnosis of MCI \cite{48, 49}.

2.3 Study participants

The initial participant selection criteria were as follows: 1) available CDR score associated with the time of image acquisition, and 2) available 3T T1 scanning protocol to ensure maximum resolution for the image analysis.

We found 204 unique participants in ADNI-1 with available 3T T1 MR images. Image data were available for all participants at different time points ranging from baseline to month 24. Since we were interested in predicting static cognition levels (CDR scores, cognitive groups), the time point was irrelevant. We selected one time point per participant to ensure unique participants across groups. To maximize group sizes, we first selected participants in the minority group, CDR score 2. These participants were excluded from all the other groups. Participants with CDR scores 1 and 0.5 were selected next. All the remaining participants not assigned to any groups were placed in the CDR 0 group. CDR score 3 was excluded due to small sample size. Next, we proceeded to find the 3T MR scan time points associated with the assigned group labels for participants. The image data acquired at the selected time points were used for analysis. 31 participants in total were excluded. The exclusions were either due to a mismatch between imaging and CDR score acquisition date (n=21), or image unavailability (n=10). This lead to a final sample size of 173: 67 non-impaired (CDR 0), 48 questionable (CDR 0.5), 39 mild (CDR 1), and 19 moderate (CDR 2) cognitively impaired individuals.

Table 1 describes demographic and clinical characteristics of the included study participants. Note that to receive a diagnosis of MCI or AD, in addition to clinician judgement, intra-individual decline must be obtained with serial cognitive measurements (multiple CDR scores over time), or by a history of change from previously attained levels \cite{50}. Thus, the numbers of participants between cognitive grouping and CDR scores differs.

2.4 Image preprocessing

MR images can have large intensity variations when acquired from different scanners or under different acquisition parameters. ADNI performs several preprocessing steps on magnetization-prepared rapid gradient-echo (MP-RAGE) sequence images. This includes gradwarp geometry distortion correction, B1 and N3 intensity non-uniformity corrections \cite{51} to ensure comparability of images across devices and protocols. To ascertain the comparability of images across patients, we normalized all images to have a common mean
and variance in cerebrospinal fluid (CSF) \cite{52}. Texture and volume analyses were performed using the normalized images.

### 2.5 Texture Analysis

The imaging data were imported into the MIPAV (Medical Image Processing, Analysis, and Visualization) application version 7.2.0 \cite{53}. To avoid resampling the images, we limited the segmentation of the hippocampus to the coronal view since it provided a common pixel spacing of \((1.02, 1.02)\)mm across all patients. An expert identified 3 slices with the largest possible view of bilateral hippocampi and manually placed rectangular ROIs \((16\times16 \text{ pixels})\) on the hippocampi area while avoiding inclusion of areas outside the hippocampus (Figure 1a) as much as possible. This segmentation process resulted in 6 ROIs \((3 \text{ slices } \times 2 \text{ hippocampi})\) per patient. This segmentation is considered greater than 2D and less than 3D (often referred to as 2.5D), and improves the reliability of the sampling process. The ROIs were cropped out of the images and set aside for texture analysis. Individuals manually placing ROIs on the hippocampi were blinded to the diagnosis and another blinded individual performed quality control checks to ensure ROIs were centrally placed.

Next, we acquired mean, standard deviation, and range of voxel intensities across the ROIs (subsequently referred to as Raw intensity features). We then mapped the dynamic ranges of intensities inside the ROIs to 0–255 as a preprocessing step for characterization of texture. Several statistical and spectral texture analysis methods are included in our radiomics pipeline. Textural features describing patterns or spatial distribution of voxel intensities were calculated from second-order statistical Gray Level Co-occurrence Matrices (GLCM)\cite{54}, Laplacian of Gaussian Histogram (LoGHist)\cite{55}, rotationally invariant Discrete Orthonormal Stockwell Transform (DOST)\cite{56}, Gabor Filter Banks (GFB)\cite{57}, and Local Binary Patterns (LBP)\cite{58}. These methods were implemented in Python programming language using custom-written code and open source libraries \cite{59, 60}. In total, we extracted 119 features per ROI: 3 Raw intensity, 26 GLCM, 10 DOST, 36 LoGHist, 12 LBP, and 32 GFB features. Extensive details on these features can be found in \cite{42, 43, 61}. To account for sampling variability, we averaged the features over slice without losing the laterality information, leading to a total of 238 texture features \((119 \text{ per hippocampus})\) per patient.

### 2.6 Volumetric features

We used an available online framework for computation of hippocampal volumetric measurements. Given a stack of MR images, volBrain \cite{62, 63} automatically segments parenchyma, brain tissues, macrostructure and subcortical structures (shown in Figure 1b) and reports volumetric measurements of the structures. For this study, we used 2 volumetric features for the hippocampus area including relative volume \(\%\) and asymmetry index \(\%\). Relative volume represents the sum of hippocampi volumes in relation to the volume of intracranial cavity. The asymmetry index is the difference between right and left volumes divided by their mean.
2.7 Statistical analysis and machine learning

Age and sex differences between groups were tested using the Student’s t-test and Pearson’s chi-square test, respectively. Statistical significance level was defined as p<0.05. We performed univariate analysis to compare the difference in texture and volume feature values for both CDR groups and cognitive groups. The p-values were adjusted for multiple comparisons using the Benjamini and Hochberg False Discovery Rate (FDR) method [64].

We applied Principal Component Analysis (PCA) to reduce dimensionality of texture features [65]. To maintain interpretability of the principal components, PCA was applied to features stemming from a common texture analysis method. Several comparative datasets were generated with PCA to find the optimal level of variance. The final set of PCs represented 90% of the variance in the original features. Texture PCs combined with volume features were used in supervised classification of 2 label variables: (1) cognitive groups (CN, MCI, AD), and (2) CDR scores.

Machine learning was conducted utilizing the open-source python-based package scikit-learn [66] and custom-written scripts. We used a leave one out cross validation (LOOCV) scheme to predict the labels [65] and to select features for training. LOOCV iteratively uses all samples except for one for model training. In each round, the left-out sample serves as the test case to assess the generalizability of the trained model on an unseen case. In each round, a trained model was generated using features selected by Sequential Forward Feature Selection (SFFS) [65] scheme and an internal cross-validation (CV). Starting from an empty set, SFFS sequentially added features as long as their addition resulted in CV accuracy improvement of 5%. We used Diagonal Quadratic Discriminant Analysis (DQDA) as the classification method [65]. DQDA is a naïve Bayes classifier that allows for diagonal class covariance matrices and has shown to be successful in classification tasks of high dimensional data with small sample sizes [67]. Several studies have shown that DQDA has comparable or better performance than support vector machine (SVM) in classification of high-dimensional data [68, 69].

Our data, by its nature, contained class imbalance, in which dominance of the majority class can hinder the classifier’s ability to learn the inherent properties of each class. To ensure generalizability of the result in experiments with substantial class imbalance, we used an ensemble down-sampling approach coupled with the above-mentioned learning scheme. In each cross-validation round the training samples were divided into majority and minority groups. The majority group was then randomly divided into subsets roughly the same size as the minority group. Each of the subsets were merged with the minority group and served as the training set. The average probability across models for the test sample was used as the probability for said sample. This iterative process allowed every sample in the data served as the left out sample once.

The area under the receiver operating characteristic curve (AUC-ROC), sensitivity, and specificity were used to assess classification performance using the open-source software packages R (2.7) [70] and Scipy (0.15.1, Python 2.7) [71]. The method by DeLong et al. and the pRoc package [72] were used to estimate ROC curve significance, p-values, and 95% confidence intervals [73]. The significance level (p<0.05) is the probability that the
observed sample area under the ROC curve is significantly different from the null hypothesis (Area = 0.5), and is evidence that the model does have an ability to distinguish between the two groups.

3. Results

The MCI group had a higher proportion of males than the CN and AD groups (Pearson chi-square = 5.2120, p=0.02). No significant difference was observed in sex ratio of the other groups. Including sex in models with texture did not impact results. As expected, the age of participants in the CDR 2 group was significantly higher than other CDR levels. Including age in models with volume did not impact results. Figure 2 compares volume features across groups and CDR scores. Figure 3 shows the univariate comparison of features across feature groups. Features extracted from left and right hippocampi showed similar significance levels. Increasing the level of variance included in the principal components of texture features did not improve the results.

3.1 Prediction of Cognitive Groups

Figure 4A shows the area under the ROC curves (AUCs) for the cognitive groups. Classification reached AUC levels of: 0.89 (CI: 0.82–0.94) for CN – AD; 0.86 (CI: 0.79–0.91) for CN – MCI; and, 0.70 (CI: 0.61–0.77) for MCI-AD. Table 2 shows the performance measures, selected features, and ROC curve analysis for the cognitive groups. All three models were significant at p ≤0.05. Including sex in the models did not impact results.

3.2 Prediction of Clinical Dementia Rating Scores

Figure 4B shows the area under the ROC curves (AUCs) for CDR groups. The AUC levels of our models were: 0.98 (CI: 0.93–0.99) for CDRs 0–2; 0.95(CI: 0.9–0.98) for CDRs 0–1; 0.84 (CI: 0.76–0.89) for CDRs 0–0.5; 0.73 (CI: 0.61–0.83) for CDRs 0.5–2; 0.71 (CI: 0.61–0.8) for CDRs 0.5–1; and, 0.56 (CI: 0.42–0.69) for CDRs 1–2. Overall, models were more successful in classification when the target groups were further apart on the CDR spectrum. Table 3 presents details of the models’ performance and significance, selected features, and ROC curve statistics for this section. All models were significant at p ≤0.05 except for the classification model CDR 1–2. Relative volume of hippocampi (%volume) was a predictive feature in two of the six models. We conducted further analysis to assess whether age accounted for the significance of %volume. When age was included in the model, %volume remained highly statistically significant (p=0.003) while age was not significant (p=0.35). The AUC only slightly increased from 0.98 (model with %volume alone) to 0.9910 (model with %volume and age). A model containing age by itself resulted in an AUC of only 0.785, and the addition of %volume significantly improved the model fit (p<0.0001). Thus, we conclude that %volume is meaningful in differentiating between CDR 0 and 2, independent of age.

4. Discussion

Here we report that the well-established MR volume features and radiomics texture features had comparable and complimentary utility in classifying cognitive groups and CDR.
There is ample literature on the utility of imaging features extracted from MRI to assist in clinical diagnosis of probable AD. Several investigations have focused on using volume, shape, and other structural MR features in identifying CN, MCI, and AD \cite{10,13,18,26,28,30,74-78}. Texture features have also been used in identifying AD \cite{14,28,32-35,79}. The literature is controversial about what exactly texture captures in the context of AD. Sorensen et al. speculate that texture patterns may provide information on hippocampal function, due to the significant correlation with FDG-PET uptake \cite{14}. The same group also found that hippocampal texture, followed by hippocampal volume, were the most significant features in their algorithm to discriminate cognitive groups \cite{35}.

Our results are consistent with that of Sorensen et al 2016 \cite{14}. For example, when they only used volume to discriminate between ADNI CN-AD, they achieved an AUC of 0.91. In our case, we achieved an AUC of 0.89 on this task. They also used texture features to differentiate CN and MCI with an AUC of 0.76, comparable to our AUC of 0.86 for the same task.

One technical difference between our methods and those of Sorensen et al. was that they resampled MR images to have consistency in image voxel size across their cohort. Resampling is often a necessary preprocessing step when images are obtained using different imaging protocols or devices. However, resampling involves interpolation, which can affect the spatial frequency content of the image. In order to establish a reliable baseline for the utility of texture features, we focused on images with a common voxel size in this study. We also used 3T imaging for higher spatial resolution and contrast-to-noise ratios. Another difference between Sorensen’s work and ours is that we used texture features to predict CDR scores. We were able to distinguish CDR 0 (no impairment) from 1 (mild dementia) with an AUC of 0.95. This model used a variety of texture features, but not hippocampal volume. On the other hand, volume features alone were able to distinguish CDR 0 from 0.5 (questionable impairment) with an AUC of 0.84. They also were able to distinguish CDR 0 from 2 (moderate dementia) with an AUC of 0.98. Overall, our CDR models performed well at distinguishing cognitively normal people from those with the early stage of, or questionable, cognitive impairment.

Distinguishing between CDR 1 and 2 was the most difficult task in our study, and AUC classification performance was poor, not achieving statistical significance (p=0.46). The transition from mild to moderate impairment appears to be a subtle shift without pronounced, discernable changes in texture or hippocampal volume. While texture features suggest that CDR scores and neuropathology may have a relationship early in cognitive impairment, i.e. early deposition of amyloid or tau, the lack of discrimination accuracy between CDR 1 and 2 suggests the pathological depositions may not help in improving classification accuracy. Aisen et al posit that the terminology behind mild and moderate AD is inaccurate, because the individual has had the disease present for many years \cite{80}. The clinical staging nomenclature infers a clear distinction between various stages, but in reality, the process progresses in a more continuous manner \cite{80}.

Due to technical limitations of our pipeline, we did not perform 3D segmentation of the hippocampi. Instead, we used a 2.5D segmentation approach in which the hippocampi were
segmented on several 2D slices to increase texture sampling. In this approach, we manually placed 2D ROIs on three slices with the largest cross-sectional view of the hippocampus (16×16 pixels). We acknowledge that extracted ROIs may have potentially included immediate anatomical structures such as the entorhinal cortex, resulting in mixed captured signals. In future studies, we plan to replicate the study using an automatic segmentation process.

Small sample size is another limitation of this study (N=173). When divided between CDR groups, each dataset consisted of few samples with a high-dimensional feature space, two known contributors to model overfitting. Due to the lack of sufficient sample size, we did not split the dataset into train and test sets. In order to provide a realistic estimation of model performance and avoid overfitting, we adopted a nested cross-validation scheme for model training and validation and a rather conservative threshold for feature selection (minimum of 5% CV accuracy improvement). Given that our results are comparable with previous studies, we feel confident that the risk of overfitting was mitigated and that the results presented here are generalizable to external data. In the future, we aim to validate this result on larger external datasets. Lastly, the reader should note that we cannot claim the clinical utility of textural biomarkers introduced here since the models were not tested prospectively.

5. Conclusion

We utilized existing resources (ADNI-1 data) to introduce a new application of brain MR radiomics using texture analysis and volumetric features in the field of aging, neuropsychiatry, and dementia. Our study findings support the use of brain MR radiomics features for identifying early cognitive impairment, as many features are sensitive to early AD pathology. Future studies need to replicate these findings and should examine the clinical utility of MR texture features as AD biomarkers. Beyond volume and texture analysis of T1 images of the hippocampus, future applications should expand to incorporate additional data sources. These could include additional MRI contrasts (for example, diffusion tensor imaging), fMRI, and PET. Additional brain structures, known to be involved in AD progression, could also be investigated.

Acknowledgments

This publication was made possible by CTSA Grant Number UL1 TR000135 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH. 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.; MesoScale 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...
References


46. ADNI Study Design; Background & Rationale. 4–6-2018]; Available from: http://adni.loni.usc.edu/study-design/background-rationale/.


51. MRI Pre-processing: Image Corrections Provided by ADNI. 11–08-16]; Available from: http://adni.loni.usc.edu/methods/mri-analysis/mri-pre-processing/.


FIGURE 1. Segmentation of the hippocampus in texture and volume analysis
(a) Texture analysis regions of interest: The left and right hippocampal areas are manually marked using 16×16 pixel squares (contour in red). This process is repeated on 3 coronal slices with the largest crosssectional view of the hippocampus area. (b) Volume analysis region of interest: volBrain pipeline segments subcortical brain tissues and reports their volumetric measurements. The image shows overlay of volBrain segmentation results. Hippocampal areas, the region of interest in our analyses, are shown in orange.
FIGURE 2. Comparison volume features across cognitive groups and CDR scores
The plot shows the distribution of the two volume features (y-axis) across different grouping of participants: cognitive states and CDR scores (x-axis). %Volume shows the sum of hippocampal volumes in relation to the volume of intracranial cavity. The asymmetry index shows the difference between right and left hippocampal volumes divided by their mean.
[Key: CN: cognitively normal; MCI: mild cognitive impairment; AD, Alzheimer’s disease; CDR: clinical dementia rating]
FIGURE 3. Radiomic features that help differentiate CDR scores and cognitive groups
Dependent variables are listed above columns (CDR score and cognitive group). We separated the data into different combinations of binary scores for each dependent variable and performed univariate analysis. Color maps show the False Discovery Rate (FDR) corrected p-values of a two-sample t-test within the dataset of each classification problem. Red to white colors indicates significant (low) p-values. A lower p-value indicates a better ability to differentiate the pair of dependent variables in the column title. [Key: Dost: Discrete Orthonormal Stockwell Transform; Gabor: Gabor Filter Banks; GLCM: Gray level]
Co-occurrence Matrices; LBP: Local Binary Patterns; LoGHist: Laplacian of Gaussian Histograms; HC: Hippocampus]
FIGURE 4. Comparison of Area Under ROC Curves
A) Area under the ROC curves of cognitive group classification models; B) Area under the receiver operator curve of CDR score classification models. Error bars show the confidence interval of the AUCs. [Key: CN: cognitively normal; MCI: mild cognitive impairment; AD, Alzheimer’s disease; CDR: clinical dementia rating; ROC: receiver operating characteristic curve; AUC: area under curve.]
## TABLE 1.

Demographic and Clinical Characteristics of the Samples

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cognitively Normal (CN) N=62</th>
<th>Mild Cognitive Impairment (MCI) N=70</th>
<th>Alzheimer’s disease (AD) N=41</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Age</td>
<td>75.2 (4.7)</td>
<td>76.0 (8.4)</td>
<td>76.1 (8.7)</td>
</tr>
<tr>
<td>Sex, Male</td>
<td>26 (41.9)</td>
<td>43 (%)</td>
<td>61.4 †</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CDR 0 N=67</th>
<th>CDR 0.5 N=48</th>
<th>CDR 1 N=39</th>
<th>CDR 2 N=19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Age</td>
<td>74.9 (5.3)</td>
<td>75.8 (7.3)</td>
<td>74.3 (8.9)</td>
<td>81.3 (7.6)</td>
</tr>
<tr>
<td>Sex, Male</td>
<td>29 (43.3)</td>
<td>25 (52.1)</td>
<td>22 (56.4)</td>
<td>9 (47.4)</td>
</tr>
</tbody>
</table>

[Key: CDR: clinical dementia rating; AD: Alzheimer’s disease; MCI: mild cognitive impairment; CN: cognitively normal. ]% refers to percentage of the specific group, SD is standard deviation.

†indicates significantly higher proportion of males in the MCI group (pearson chi-square= 5.2120, p=0.02), and

*indicates a significantly higher age in the CDR 2 group (student’s t-test, p<0.0001).
**TABLE 2.**

Classification Results for Cognitive Groups

<table>
<thead>
<tr>
<th>Cognitive Groups</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Feature Type</th>
<th>Features</th>
<th>Standard Error(\text{[73]})</th>
<th>95% Confidence Interval</th>
<th>Z statistic</th>
<th>Significance level P (Area&lt;0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN-MCI</td>
<td>0.86</td>
<td>0.79</td>
<td>0.83</td>
<td>Texture</td>
<td>Left HC LoGHist pc 1 Right HC LBP pc 1</td>
<td>0.03</td>
<td>(0.79, 0.91)</td>
<td>11.58</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MCI-AD</td>
<td>0.70</td>
<td>0.54</td>
<td>0.83</td>
<td>Texture</td>
<td>Left HC LBP pc 1</td>
<td>0.05</td>
<td>(0.61, 0.77)</td>
<td>4.16</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>CN-AD</td>
<td>0.99</td>
<td>0.82</td>
<td>0.87</td>
<td>Volume</td>
<td>% HC Volume</td>
<td>0.03</td>
<td>(0.82, 0.94)</td>
<td>12.31</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Key: CN: cognitively normal; MCI: mild cognitive impairment; AD: Alzheimer’s disease; LoGHist: Laplacian of Gaussian Histograms; %Volume: relative volume in percent; LBP: Local Binary Patterns; HC: Hippocampus; PC: Principal Component.
### TABLE 3.
Classification Result for Prediction of Clinical Dementia Rating (CDR) Score

<table>
<thead>
<tr>
<th>CDR Pairs</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Feature Type</th>
<th>Features</th>
<th>Standard Error (73)</th>
<th>95% Confidence Interval</th>
<th>Z statistic</th>
<th>Significance level P (Area = 0.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0, 0.5</td>
<td>0.84</td>
<td>0.78</td>
<td>0.81</td>
<td>Volume</td>
<td>% HC Volume</td>
<td>0.04</td>
<td>(0.76, 0.89)</td>
<td>9.67</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>0.5, 1</td>
<td>0.71</td>
<td>0.77</td>
<td>0.67</td>
<td>Texture</td>
<td>Right HC Dost pc2</td>
<td>0.05</td>
<td>(0.61, 0.8)</td>
<td>4.03</td>
<td>0.0001</td>
</tr>
<tr>
<td>1, 2</td>
<td>0.56</td>
<td>0.58</td>
<td>0.59</td>
<td>Texture</td>
<td>Left HC Gabor pc 1</td>
<td>0.08</td>
<td>(0.42, 0.69)</td>
<td>0.74</td>
<td>0.46</td>
</tr>
<tr>
<td>0, 1</td>
<td>0.95</td>
<td>0.88</td>
<td>0.96</td>
<td>Texture</td>
<td>Left HC Dost pc1, Left HC LoGHist pc5, Left HC Gabor pc1, Right HC GLCM pc2</td>
<td>0.02</td>
<td>(0.9, 0.98)</td>
<td>22.88</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>0.5, 2</td>
<td>0.73</td>
<td>0.58</td>
<td>0.90</td>
<td>Texture</td>
<td>Left HC Gabor pc1, Left HC Dost pc1</td>
<td>0.08</td>
<td>(0.61, 0.83)</td>
<td>2.89</td>
<td>0.0038</td>
</tr>
<tr>
<td>0, 2</td>
<td>0.98</td>
<td>1.0</td>
<td>0.90</td>
<td>Volume</td>
<td>% HC Volume</td>
<td>0.01</td>
<td>(0.93, 0.99)</td>
<td>46.5</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Key: CDR: Clinical Dementia Rating; Gabor: Gabor Filter Banks; Dost: Discrete Orthonormal Stockwell Transform; % Volume: relative volume in percent; LBP: Local Binary Patterns; HC: Hippocampus; LoGHist: Laplacian of Gaussian Histograms; GLCM: Gray level Concurrence Matrices; PC: Principal Component.