

1 **Deep learning reveals pathology-confirmed neuroimaging** 2 **signatures in Alzheimer’s, vascular and Lewy body** 3 **dementias**

4 Di Wang,¹ Nicolas Honnorat,¹ Jon B. Toledo,² Karl Li,¹ Sokratis Charisis,¹ Tanweer Rashid,¹
5 Anoop Benet Nirmala,¹ Sachintha Ransara Brandigampala,¹ Mariam Mojtabai,¹ Sudha Seshadri,¹
6 the Alzheimer's Disease Neuroimaging Initiative and Mohamad Habes^{1,3}

7 **Abstract**

8 Concurrent neurodegenerative and vascular pathologies pose a diagnostic challenge in the clinical
9 setting, with histopathology remaining the definitive modality for dementia-type diagnosis. To
10 address this clinical challenge, we introduce a neuropathology-based, data-driven, multi-label deep
11 learning framework to identify and quantify *in-vivo* biomarkers for Alzheimer's disease (AD),
12 vascular dementia (VD), and Lewy body dementia (LBD) using antemortem T1-weighted MRI
13 scans of 423 demented and 361 control participants from NACC and ADNI datasets. Based on the
14 best-performing deep learning model, explainable heatmaps are extracted to visualize disease
15 patterns, and the novel **Deep Signature of Pathology Atrophy REcognition** (DeepSPARE) indices
16 are developed, where a higher DeepSPARE score indicates more brain alterations associated with
17 that specific pathology.

18 A substantial discrepancy in clinical and neuropathology diagnosis was observed in the demented
19 patients: 71% of them had more than one pathology, but 67% of them were clinically diagnosed
20 as AD only. Based on these neuropathology diagnoses and leveraging cross-validation principles,
21 the deep learning model achieved the best performance with a balanced accuracy of 0.844, 0.839,
22 and 0.623 for AD, VD, and LBD, respectively, and was used to generate the explainable deep-
23 learning heatmaps and DeepSPARE indices. The explainable deep-learning heatmaps revealed
24 distinct neuroimaging brain alteration patterns for each pathology: the AD heatmap highlighted
25 bilateral hippocampal regions, the VD heatmap emphasized white matter regions, and the LBD
26 heatmap exposed occipital alterations. The DeepSPARE indices were validated by examining their

© The Author(s) 2024. Published by Oxford University Press on behalf of the Guarantors of Brain. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

1 associations with cognitive testing, neuropathological, and neuroimaging measures using linear
2 mixed-effects models. The DeepSPARE-AD index was associated with MMSE, Trail B, memory,
3 hippocampal volume, Braak stages, CERAD scores, and Thal phases ($p_{FDR-adjusted} < 0.05$). The
4 DeepSPARE-VD index was associated with white matter hyperintensity volume and cerebral
5 amyloid angiopathy ($p_{FDR-adjusted} < 0.001$). The DeepSPARE-LBD index was associated with
6 Lewy body stages ($p_{FDR-adjusted} < 0.05$). The findings were replicated in an out-of-sample ADNI
7 dataset by testing associations with cognitive, imaging, plasma, and CSF measures. CSF and
8 plasma pTau181 were significantly associated with DeepSPARE-AD in the AD/MCI $_{A\beta+}$ group
9 ($p_{FDR-adjusted} < 0.001$), and CSF α -synuclein was associated solely with DeepSPARE-LBD
10 ($p_{FDR-adjusted} = 0.036$).

11 Overall, these findings demonstrate the advantages of our innovative deep-learning framework in
12 detecting antemortem neuroimaging signatures linked to different pathologies. The newly deep
13 learning-derived DeepSPARE indices are precise, pathology-sensitive, and single-valued
14 noninvasive neuroimaging metrics, bridging the traditional widely available *in-vivo* T1 imaging
15 with histopathology.

16 **Author affiliations:**

17 1 Neuroimage Analytics Laboratory and Biggs Institute Neuroimaging Core, Glenn Biggs Institute
18 for Neurodegenerative Disorders, University of Texas Health Science Center at San Antonio, San
19 Antonio, Texas 78229, USA

20 2 Nantz National Alzheimer Center, Stanley Appel Department of Neurology, Houston Methodist
21 Hospital, Houston, Texas 77030, USA

22 3 Biomedical Image Analytics Division, Research Imaging Institute, University of Texas Health
23 Science Center at San Antonio, San Antonio, Texas, USA

24 Correspondence to: Mohamad Habes

25 8403 Floyd Curl Dr, San Antonio, TX 78229, USA

26 E-mail: habes@uthscsa.edu

27 **Keywords:** deep learning; Alzheimer's disease; vascular dementia; Lewy body dementia; MRI;
28 neuropathology

1 **Running Title:** AI reveals mixed-dementia patterns

2

3 **Introduction**

4 Aging is associated with the accumulation of neurodegenerative and vascular pathologies in the
5 brain, which can lead to the development of cognitive impairment and dementia¹. In 2024, an
6 estimated 6.9 million Americans aged 65 and older live with Alzheimer's disease (AD), the most
7 common cause of dementia, with vascular dementia (VD) and dementia with Lewy bodies (LBD)
8 following^{2,3}.

9 Clinical dementia diagnosis work-up relies on history taking, physical exam, cognitive testing,
10 blood work, and neuroimaging biomarkers⁴⁻⁶. Neuroimaging biomarkers are particularly
11 interesting due to the range of modalities and sequences available. MRI scans are usually included
12 as a part of the initial clinical assessment of dementia, providing the opportunity to detect and
13 quantify different neurodegenerative pathologies⁷⁻¹⁰. MRI has been most frequently used for
14 detecting and quantifying AD-related hippocampus atrophy^{7,11-13}, as well as different
15 endophenotypes of vascular brain injury, such as white matter hyperintensities (WMHs)^{1,14,15},
16 hemorrhages¹⁶, infarcts¹⁷, and enlarged perivascular spaces¹⁸. For LBD, MRI scans usually present
17 preserved medial temporal lobe volume¹⁹ and occipital atrophy²⁰⁻²². Other neuroimaging
18 biomarkers, such as amyloid- β ($A\beta$), tauopathy captured with PET scans, and α -synuclein
19 detected by seed amplification assays (SAA) for Lewy body pathology^{23,24}, can also support the
20 diagnosis^{25,26}. However, those are not usually included in the initial assessment and thus have
21 limited clinical availability. Therefore, many MRI-based neuroimaging studies have been designed
22 to use machine learning models to predict clinical dementia diagnosis^{7,27-29}.

23 However, clinical diagnosis faces multiple challenges regarding diagnosis accuracy³⁰⁻³², where
24 neuropathology is considered the diagnostic gold standard³³⁻³⁶ because neurodegenerative diseases
25 are histopathologically characterized by the deposition of misfolded proteins with different
26 spreading patterns and underlying biological mechanisms. Thus, machine learning-based
27 biomarker development using clinical rather than neuropathological diagnosis may lead to
28 misestimation, inaccurate cutoffs³⁶, and limited accuracy^{8,37-40}.

29 Recently, the emergence of deep learning has significantly transformed the traditional machine
30 learning approach⁴¹. With their flexible architectures, deep learning models require only minimal

1 pre-processing while reaching state-of-the-art diagnostic accuracy⁴². However, most deep
2 learning-based biomarker developments⁴¹ still leverage clinical diagnosis, potentially limiting
3 their accuracy in predicting histopathology outcomes. Moreover, recent work has shown that deep
4 neural network decisions are biologically informed, as evidenced by hippocampal-focused
5 explainable heatmaps in AD classification tasks^{43–45}, which is not the case for the traditional
6 machine learning models⁴⁵. Most previous studies have focused on binary tasks. This work could
7 be among the first attempts to extract explainable heatmaps for multiple co-pathologies from a
8 single multi-label model.

9 Previously, we have developed a set of machine learning-derived SPARE (Spatial Pattern of
10 Atrophy for Recognition) single-valued metrics that capture distinct dimensions of brain
11 alterations linked to cognitive decline⁴⁶. For instance, SPARE-AD⁴⁷ quantifies AD-like atrophy,
12 while SPARE-BA⁴⁸ measures atrophy related to brain aging. However, these indices relied on
13 traditional support vector machine (SVM) methods and did not consider the possibility of co-
14 pathologies^{30–32,49}. Postmortem studies showed that most dementia cases are heterogeneous with
15 multiple co-pathologies³⁰. This underscores the need for more sophisticated methods to
16 disentangle dementia heterogeneity⁵⁰ using *in-vivo* imaging.

17 In this study, we develop a multi-label deep learning network trained with neuropathologically
18 confirmed diagnoses to quantify how atrophy relates to the three most common co-pathologies.
19 We utilize a combined dataset from two large neuroimaging repositories: the Alzheimer's Disease
20 Neuroimaging Initiative (ADNI)³⁷ and the National Alzheimer's Coordinating Center (NACC)⁵¹.
21 Specifically, our models are trained to detect the presence of AD, VD, and LBD pathologies in a
22 total of 423 demented and 361 control participants using their antemortem T1-weighted MRI
23 scans. In addition, we benchmark our deep learning model against traditional machine learning
24 techniques by training multi-label random forest (RF)⁵² and linear SVM⁵³ models for comparative
25 analysis. Then, we generate explainable deep-learning heatmaps to reveal the brain changes
26 detected by the model and develop DeepSPARE indices, the advanced indices specifically
27 designed to detect brain changes associated with multiple co-pathologies: AD (DeepSPARE-AD),
28 VD (DeepSPARE-VD), and LBD (DeepSPARE-LBD). The DeepSPARE indices are further
29 validated using a set of cognitive, neuropathology, and neuroimaging measures, and our findings
30 are replicated on an out-of-sample cohort of ADNI participants.

1 **Materials and methods**

2 **Neuroimaging data and neuropathology diagnosis criteria**

3 The machine learning models were trained with 3D T1-weighted MRI scans provided by NACC⁵¹
4 and ADNI³⁷. For each study participant, the most recent MRI scan was selected to optimize the
5 proximity between these antemortem scans and the neuropathological diagnosis obtained at the
6 time of death. A total of 784 participants were included in this study to derive our machine learning
7 models: 423 neuropathology-identified demented participants and 361 cognitively healthy
8 controls. The 423 demented participants were identified as having AD, VD, or LBD pathology
9 based on the following criteria. Participants had AD pathology if their AD neuropathological
10 change scores, a composite score of Braak staging⁵⁴, Thal phase⁵⁵, and CERAD score⁵⁶, reflected
11 moderate or high AD neuropathological change⁵⁷. Participants who had VD pathology were
12 selected by the presence of moderate or severe arteriosclerosis, cerebral amyloid angiopathy
13 (CAA), or infarcts and hemorrhages⁵⁸. LBD pathology participants were selected based on limbic
14 or neocortical Lewy body pathology presence¹⁹. An out-of-sample ADNI dataset was included in
15 the study to validate our findings, consisting of 1041 participants diagnosed with AD, mild
16 cognitive impairment (MCI), and controls. **Table 1** describes their demographical characteristics.

17 **Data processing**

18 A fully automatic image segmentation pipeline, the multi-atlas region segmentation utilizing
19 ensembles pipeline, was used to correct field bias and skull-strip all the T1-weighted brain scans⁵⁹.
20 The skull-stripped scans were then registered to the 1 cubic millimeter resolution 2009c version of
21 the ICBM152 MNI atlas⁶⁰ of size 193x229x193 using the non-rigid SyN registration method⁶¹
22 provided with the ANTs library (version 2.3.4)⁶². Then, all scans were downsampled to a size of
23 65x77x65 using the Nilearn Python package *resample_img*. This downsampling was introduced
24 to reduce computational time, allowing for the exploration of a broad set of statistical models and
25 the conducting of extended cross-validation and replication experiments. The downsampling has
26 also produced smaller deep networks, usually associated with lower risks of overfitting⁶³ and
27 improved accuracy⁶⁴. After downsampling, each scan intensity was normalized using the min-max
28 normalization approach.

1 Multi-label classification models

2 Three machine learning models were trained to predict the neuropathology diagnosis from the
 3 processed 3D T1-weighted MRI scans: RF models, ensemble linear SVM models, and deep neural
 4 networks based on a 3D ResNet architecture⁶⁵. In a multi-label classification task, the model aims
 5 at predicting more than one mutually non-exclusive label. This framework allows each input to be
 6 associated with multiple labels simultaneously. Here, the model must predict three labels for each
 7 subject, representing the likelihood of AD, VD, and LBD. This approach allows the model to
 8 independently identify the presence or absence of each condition, providing a comprehensive
 9 assessment of the subject's potential pathologies. A detailed explanation can be found in
 10 **Supplementary S2**. A multi-label asymmetric loss was adopted to mitigate the effects of class
 11 imbalance for multi-label deep neural networks⁶⁶. This asymmetric loss exhibits two advantages:
 12 it reduces the impact of imbalanced classes during binary classification, and it reduces the
 13 influence of easily classified negative samples⁶⁶. The asymmetric loss was compared to a standard
 14 binary cross-entropy loss. A detailed explanation can be found in the **Supplementary S2**. Because
 15 a multi-label classification cannot be directly conducted using a single SVM, an ensemble of three
 16 linear SVM binary classifiers was trained, one per pathology, to accomplish each multi-label task.
 17 The code from this study is available at <https://github.com/UTHSCSA-NAL/DeepSPARE>.

18 Model evaluation

19 The classification performance was tenfold cross-validated and evaluated by the following
 20 classification metrics: a multi-label precision, a multi-label recall, a multi-label F1, a balanced
 21 accuracy (*BACC*) for each dementia pathology, an average balanced accuracy (*BACC_{avg}*), and the
 22 area under the ROC curve (AUC) for each dementia pathology. These metrics are defined below,
 23 and a more detailed explanation of those metrics can be found in the **Supplementary S3**. Here,
 24 TP and TN denote true positive and true negative values, and FP and FN denote false-positive and
 25 false-negative values. *BACC_{AD}*, *BACC_{VD}*, and *BACC_{LBD}* denote the balanced accuracy for AD,
 26 VD, and LBD pathology, respectively.

$$27 \text{ precision} = \frac{TP_{AD} + TP_{VD} + TP_{LBD}}{TP_{AD} + TP_{VD} + TP_{LBD} + FP_{AD} + FP_{VD} + FP_{LBD}} \quad (1)$$

$$28 \text{ recall} = \frac{TP_{AD} + TP_{VD} + TP_{LBD}}{TP_{AD} + TP_{VD} + TP_{LBD} + FN_{AD} + FN_{VD} + FN_{LBD}} \quad (2)$$

$$29 \text{ F1} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

$$1 \quad BACC = \frac{1}{2} \left[\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right] \quad (4)$$

$$2 \quad BACC_{avg} = \frac{1}{3} [BACC_{AD} + BACC_{VD} + BACC_{LBD}] \quad (5)$$

3 For the three types of multi-label models, the model achieving the best F1 and $BACC_{avg}$ among
4 tested parameters was retained to derive ROC curves and further analyses. The AUC of those three
5 best models was compared using the Delong test⁶⁷.

6 **Model interpretation**

7 The machine learning models were first interpreted by deriving pathology probability charts
8 directly from the outputs of best RF, SVM, and deep neural networks to evaluate how well they
9 disentangle dementia pathologies. A silhouette score⁶⁸ was calculated for each machine learning
10 model to quantify its ability to separate the three pathologies. The silhouette score is a standard
11 measure of cluster separability⁶⁸, where a higher silhouette score indicates better separability
12 between given clusters. Three pathology probability charts were derived for each model: one plot
13 for each pair of pathologies, reporting the likelihood of one T1 MRI scan having the pathology
14 investigated in this study. The RF model and deep neural networks can output predicted probability
15 directly, while the SVM scores produced by the SVM models were transformed into probabilities
16 using the default cross-validated logistic regression method implemented in the scikit-learn
17 library⁶⁹.

18 Then, a set of explainable heatmaps was derived to highlight the brain regions considered by the
19 best SVM and the best deep learning model to support their predictions. SVM heatmaps were
20 generated from SVM coefficients, and each pathology heatmap was generated from the
21 corresponding binary SVM model. Based on our previous study⁴⁵, where we observed that the
22 Integrated Gradients (IG)⁷⁰ heatmaps were better at capturing brain alterations associated with AD
23 than heatmaps generated via Guided Grad-CAM⁷¹ and Layer-wise Relevance Propagation⁴⁴, the
24 IG method is used to derive both group and individual explainable deep-learning heatmaps. Group
25 deep-learning heatmaps are created by averaging the heatmaps of all subjects within the same
26 pathology group, resulting in a robust representation of brain alterations specific to each
27 pathology^{44,45}. Individual deep-learning heatmaps, derived directly from each subject, may serve
28 as a diagnostic fingerprint tool⁴⁴. Gaussian noise with a zero mean and a standard deviation of 0.1,
29 similar to the noise introduced during data augmentation, is used as the baseline for IG. All maps

1 were smoothed by applying a Gaussian kernel with a size of one voxel of standard deviation. A
2 detailed explanation of IG can be found in the **Supplementary S4**.

3 **DeepSPARE indices evaluation against cognitive testing,** 4 **neuropathology, and neuroimaging measures**

5 The DeepSPARE indices were generated by performing the logit transformation of the probability
6 of the best-performing deep learning model. Three indices were generated for each pathology:
7 DeepSPARE-AD, DeepSPARE-VD, and DeepSPARE-LBD. A set of measures was used to
8 validate further these indices: three cognitive scores, six neuropathology measures, and two
9 neuroimaging measures. Specifically, three scores capturing the global, executive, and episodic
10 memory cognitive performance were investigated: the Mini-Mental State Examination (MMSE)⁷²,
11 the Trail Making Test B (Trail B)⁷³, and the immediate logical memory subtest of the Wechsler
12 Memory Scale-Revised⁷⁴. Six neuropathology measures were examined: arteriolosclerosis⁷⁵,
13 CAA⁷⁶, Thalphase⁵⁵, Braak stage⁵⁴, CERAD⁷⁷ score, and Lewy body stages⁷⁸. Two neuroimaging
14 measures were also evaluated: the hippocampal and white matter hyperintensity (WMH) volumes.
15 Those imaging measures were performed by the IDeA Lab for the NACC dataset, following ADNI
16 protocols. Gray matter segmentation is based on an Expectation-maximization algorithm^{79,80}, and
17 WMH is calculated based on fluid-attenuated inversion recovery (FLAIR) and 3D T1 MRI⁸¹. For
18 more details, please refer to the NACC and ADNI websites.

19 The associations between the aforementioned measures and the DeepSPARE indices were
20 examined using generalized linear mixed models (GLMMs). A Gaussian conditional distribution
21 with an identity link function was specified for continuous outcomes, and a binomial conditional
22 distribution with a logit link function was specified for dichotomous outcomes. In these models,
23 the respective cognitive, neuropathology, or neuroimaging measure of interest was the outcome,
24 and the DeepSPARE indices were the main predictors (expressed as z-scores to facilitate
25 comparisons between their effect estimates). The models were adjusted for age and sex, and
26 included random intercepts for different study sites and different MRI scanners, assuming a
27 partially crossed random effects design (for more details, see **Supplementary S5**). Considering
28 the possibility of strong correlations between the different DeepSPARE indices (due to the frequent
29 occurrence of co-pathologies), we calculated their respective variance inflation factors (VIF) in
30 each model, to assess the degree of collinearity⁸². VIF values of more than 5 were considered

1 concerning for collinearity, based on more conservative literature-derived thresholds^{82,83}. False
2 discovery rate (FDR)-adjusted p -values⁸⁴ and 95% confidence intervals for coefficient⁸⁵ were
3 computed for all models. A FDR-adjusted p -value below 0.05 is considered as significant.

4 **DeepSPARE indices evaluation using out-of-sample dataset**

5 A total of 1041 ADNI participants, independent of the model training group, were used to generate
6 the pathology probability charts and validate the DeepSPARE indices. ADNI participants were
7 divided into three groups: AD and MCI amyloid- β positive (AD/MCI $_{A\beta+}$) group, MCI amyloid- β
8 negative (MCI $_{A\beta-}$) group, and control group (CN). The AD/MCI $_{A\beta+}$ group was selected based on
9 positive amyloid PET, or CSF amyloid < 977 pg/ml⁸⁶. The MCI $_{A\beta-}$ and CN groups were selected
10 based on negative amyloid PET and CSF amyloid ≥ 977 pg/ml. The total number for AD, MCI $_{A\beta+}$,
11 MCI $_{A\beta-}$, and CN groups were 314, 420, 161, and 146, respectively.

12 The DeepSPARE indices for these 1,041 ADNI participants were generated through the
13 following process. First, the ten models from the best-performing multi-label deep learning model,
14 developed during tenfold cross-validation, were applied to each of the processed T1-weighted MRI
15 scans. This results in ten results for each subject and for each pathology. Then these ten results
16 were averaged and converted into a probability for each pathology using a sigmoid transformation.
17 These probabilities were used to create pathology probability charts. Finally, the probabilities were
18 transformed into the DeepSPARE indices using the logit function.

19 For this out-of-sample dataset, the associations of DeepSPARE indices with cognitive scores,
20 neuroimaging, plasma, and CSF measures were tested with GLMMs structured similarly to those
21 described in the previous section. Those measures are MMSE, Trail B, ADAS-Cog 13⁸⁷, the
22 volume of the hippocampus and WMH, SPARE-Tau⁸⁸ (a machine learning-derived index based on
23 tau PET scans), plasma $A\beta$ -42⁸⁹, plasma tau phosphorylated at threonine-181 (pTau181)⁹⁰, the
24 ratio of CSF pTau181/ $A\beta$ 42⁹¹, CSF pTau181⁹¹, and the presence of α -synuclein in CSF by using
25 the seed amplification assays²⁴ (SAA) method provided by Amprion Clinical Laboratory^{92,93}
26 (please refer to ADNI website for more details).

1 Results

2 Clinical and neuropathology diagnoses

3 **Figure 1A** summarizes the 423 neuropathology-confirmed diagnoses of the demented participants
4 and 361 control participants included as the training set in our machine learning models. The mean
5 age of demented and control participants is 76.8, and 75.6, respectively. The number of
6 female/male participants in demented and control participants are 130/231, and 146/277,
7 respectively. There were no significant age ($p = 0.11$), sex ($p = 0.9$), or education ($p = 0.08$)
8 differences between the demented and control participants. The mean time interval between death
9 and the scan date of included MRI scans of the demented participants was 4.9 years. The majority
10 of the 423 demented participants presented multiple co-pathologies: 300 participants had at least
11 two different neurodegenerative pathologies (70.9%), including 54 participants who had all three
12 pathologies (12.8%). Only 123 participants had a single pathology (29.1%): 51 had AD (12.19%),
13 51 had VD (12.1%), and 21 had LBD (4.9%). Those three pathologies formed seven
14 neuropathology-confirmed diagnostic categories, and they were compared to the clinical diagnoses
15 released by the NACC and the ADNI consortia, as shown in **Figure 1B**. Among the 312
16 participants who had AD pathology, 242 (77.6%) had received a clinical diagnosis of AD, 52
17 (16.7%) had received an MCI clinical diagnosis, and 70 (22.4%) had received a non-AD clinical
18 diagnosis. A total of 243 (72.8%) of the 334 participants who had VD pathology had been clinically
19 diagnosed as having AD, and 3 (0.9%) had been clinically diagnosed as having VD. Of the 131
20 participants who had LBD pathology, only 23 (17.6%) had been clinically diagnosed as having
21 LBD, with most of the remaining individuals having received an AD diagnosis (75, 57.3%). These
22 counts indicate a considerable pathology overlap among dementia individuals and a substantial
23 discrepancy between clinical and neuropathology diagnoses. Tables indicating the demographics
24 of seven neuropathology-confirmed groups and the clinical diagnoses of each group are provided
25 in **Supplementary Tables 1 and 2**, respectively. The MR scanner magnetic field strength,
26 manufacturer, model name, and clinical site location used to acquire the brain scans included in
27 the machine learning model training and validation sets are reported in **Supplementary Figures**
28 **1 and 2**, respectively.

1 **Multi-label classification performance**

2 The deep neural networks trained with an asymmetric loss function and a batch size of 8 reached
3 the overall best performance, achieving the best F1, the best *BACC* for all three pathologies, and
4 the best *BACC_{avg}*, compared to all tested RF and ensemble linear SVM models. The deep learning
5 model training was completed within 3 hours for each fold (using the GPU described in
6 **Supplementary S2**).

7 **Table 2** reports the tenfold cross-validated metrics measured for the six RFs, the six ensemble
8 SVMs, and the six deep neural networks compared in this work. The training parameters of
9 asymmetric loss were selected by conducting a preliminary experiment reported in
10 **Supplementary Table 3**. The best ensemble SVM was obtained for a *c* parameter of 0.01. This
11 SVM model achieved a lower balanced accuracy than the best deep neural networks for the AD
12 and VD classification but could not classify LBD participants. It also achieved better precision but
13 a worse recall for a lower F1 than deep neural networks. Lastly, the best RF was obtained when
14 combining 300 trees, which had the worst balanced accuracies, a lower F1, and the worst recall.
15 These three models were selected to derive the ROC curves and AUC, as shown in **Figure 2A**.
16 The AUC separately generated for the three neuropathology diagnoses confirmed that the SVM
17 and deep network models were associated with similar AUC for the three diagnoses, and the RF
18 achieved the worst classification performance. More specifically, the AD AUC of 0.866 obtained
19 for the RF was significantly lower than the AUC of 0.907 obtained for the deep neural networks
20 according to the DeLong statistical test⁶⁷ ($p = 3.5 \times 10^{-7}$). The RF AUC of 0.859 measured for
21 diagnosing VD was also significantly lower than the AUC of 0.901 obtained for the deep neural
22 networks ($p = 8.3 \times 10^{-7}$, DeLong test). For LBD, RF achieved an AUC of 0.728, but the
23 difference was not significant ($p = 0.51$, DeLong test) compared to the deep neural networks' AUC
24 of 0.719.

25 **Pathology probability charts**

26 The AD, VD, and LBD probabilities directly generated by the three best machine learning models
27 were used to produce the pathology probability charts shown in **Figure 2B**. A larger silhouette
28 score was measured for the deep learning model (0.209) compared to the SVM model (silhouette
29 score = -0.162) and RF model (silhouette score = -0.193) (**Figure 2B**), indicating a better
30 separability of the pathology groups from the deep learning model. Deep learning model was the

1 only multi-label model able to disentangle all the pathologies: it performed well in AD and VD
 2 classification and reached a $BACC_{LBD}$ of 0.623 for LBD. The SVM model was able to differentiate
 3 between AD and VD but struggled to detect LBD, achieving a $BACC_{LBD}$ of 0.531. The RF model
 4 failed to distinguish the different neurodegenerative pathologies and only categorized study
 5 participants as diseased or healthy.

6 Explainable deep-learning heatmaps

7 **Figure 3** presents the group deep-learning heatmaps generated by the IG method from the best
 8 deep neural networks and the SVM coefficient maps from the best linear SVM model. Deep-
 9 learning heatmaps reflect distinct imaging patterns associated with each type of pathology, but the
 10 SVM maps do not separate the pathologies. The AD deep-learning heatmaps indicated changes in
 11 bilateral hippocampal regions. The VD heatmap focused on more white matter regions. Lastly, the
 12 heatmap derived for the LBD captured occipital regions. **Supplementary Figure 3** presents the
 13 individual heatmaps, indicating that, despite higher noise levels, the key features observed in the
 14 group heatmaps remain visible in the individual heatmaps.

15 DeepSPARE indices validation against cognitive testing, 16 neuropathology, and neuroimaging measures

17 The associations between the DeepSPARE indices and cognitive, neuropathology, and
 18 neuroimaging measures of interest are presented in **Figure 4**. All the observed associations were
 19 in their expected directions (e.g., smaller hippocampal volumes were related to higher
 20 DeepSPARE-AD values). The p -values were FDR-corrected for the eleven models reported in
 21 **Figure 4**. The VIF values indicated a relatively low degree of collinearity among the DeepSPARE
 22 indices ($VIF < 3$). The three cognitive tests were all significantly associated with DeepSPARE-AD
 23 ($p_{FDR} < 0.001$). DeepSPARE-VD was associated with the MMSE and logical memory tests (p_{FDR}
 24 < 0.001). Regarding imaging measures, the hippocampal volume was only associated with
 25 DeepSPARE-AD ($p_{FDR} < 0.001$), and the WMH volume was only associated with DeepSPARE-
 26 VD ($p_{FDR} < 0.001$). The Thal phase, CERAD score, and Braak stage were all associated with
 27 DeepSPARE-AD ($p_{FDR_{Thal}} = 0.01$, $p_{FDR_{CERAD}} < 0.001$, $p_{FDR_{Braak}} < 0.001$). CAA was only
 28 associated with DeepSPARE-VD ($p_{FDR} < 0.001$). No significant associations were observed with

1 arteriosclerosis. DeepSPARE-LBD showed no significant associations with none of these
2 measures; however, it was associated with the Lewy body stage ($p_{FDR} = 0.004$).

3 **DeepSPARE indices validation using out-of-sample dataset**

4 In the last set of experiments, the best deep neural network was applied to an out-of-sample ADNI
5 validation dataset to derive pathology probability charts (**Supplementary Figure 4**) and
6 DeepSPARE indices. The neural network predicted a significantly larger prevalence (statistical
7 results are reported in **Supplementary S7**) of dementias in the AD/MCI_{Aβ+} group, compared to
8 the MCI_{Aβ-} and the CN ADNI groups.

9 **Figures 5** and **6** present the associations of the DeepSPARE indices with cognitive and
10 neuroimaging measures in the out-of-sample ADNI dataset. The reported p -values were FDR-
11 corrected for the twenty-one models reported in the two **Figures 5** and **6**. All significant ($p_{FDR} <$
12 0.05) associations were in the expected directions for the corresponding pathology. Significant
13 associations were observed with most measures in the AD/MCI_{Aβ+} group but not in the MCI_{Aβ-}
14 group. Most of the results in the CN group were non-significant ($p_{FDR} > 0.05$) (**Supplementary**
15 **Figure 5**).

16 All cognitive scores exhibited significant associations with DeepSPARE-AD and VD indices in
17 the AD/MCI_{Aβ+} group. Notably, for MMSE, a stronger association was observed with
18 DeepSPARE-AD (estimate = -1.261, [-1.602, -0.936]) than VD (estimate = -0.573, [-0.903, -
19 0.243]). Also, for ADAS-Cog 13, a stronger association was observed with DeepSPARE-AD
20 (estimate = 5.221, [4.258, 6.193]) than VD (estimate = 1.873, [0.905, 2.843]). No significant
21 associations were observed in the MCI_{Aβ-} group.

22 A stronger association was noted between hippocampal volume and DeepSPARE-AD (estimate =
23 -0.537, [-0.652, -0.421]) than VD (estimate = -0.293, [-0.411, -0.176]). WMH volume and
24 DeepSPARE-VD were significantly associated in the AD/MCI_{Aβ+} group ($p_{FDR} < 0.001$). In the
25 MCI_{Aβ-} group, hippocampal volume was associated only with DeepSPARE-AD, and WMH was
26 associated only with DeepSPARE-VD. Additionally, SPARE-Tau was associated with
27 DeepSPARE-AD in the AD/MCI_{Aβ+} group ($p_{FDR} < 0.001$). The SPARE-Tau measure was missing
28 in most participants of the MCI_{Aβ-} group; thereby its associations with the DeepSPARE indices
29 were not examined.

1 In the AD/MCI_{Aβ+} group, CSF α -synuclein SAA was associated only with DeepSPARE-LBD
2 ($p_{FDR} = 0.036$). The CSF pTau181/A β 42 ratio and CSF pTau181 were associated only with
3 DeepSPARE-AD ($p_{FDR} < 0.001$). No significant associations between CSF measures and
4 DeepSPARE indices were observed in the MCI_{Aβ-} group.

5 Regarding plasma measures, only plasma pTau181 was significantly associated with DeepSPARE-
6 AD in the AD/MCI_{Aβ+} group ($p_{FDR} < 0.0001$). No significant associations were observed between
7 plasma measures and DeepSPARE indices in the MCI_{Aβ-} group.

8 **Discussion**

9 In this study, we developed an advanced multi-label deep-learning framework designed to
10 disentangle the three most common co-pathologies: AD, VD, and LBD, using antemortem T1-
11 weighted MRI scans. Explainable deep-learning heatmaps for these co-pathologies were generated
12 from a single model, highlighting hippocampal alterations for AD, white matter changes for VD,
13 and occipital regions for LBD. The novel neuroimaging signatures derived from our framework,
14 the DeepSPARE indices, representing the likelihood of each co-pathology, have revealed
15 significant associations with pathology-specific biomarkers and measures. Specifically, the
16 DeepSPARE-AD index was significantly associated with key indicators of AD, including
17 hippocampal volume, Braak stages, CERAD scores, and Thal phases. The DeepSPARE-VD index
18 showed significant associations with WMH and CAA. The DeepSPARE-LBD index was
19 significantly associated with Lewy body stages. The pathology sensitivity of DeepSPARE indices
20 has also been replicated in an out-of-sample dataset. This framework broadly extends the previous
21 machine learning-based neuroimaging signatures^{46–48} by enabling a more precise disentanglement
22 of the complex interplay between multiple co-pathologies through model explainable heatmaps
23 and quantitative DeepSPARE indices, serving as an advanced tool to understand and detect the
24 heterogeneity of dementia.

25 **A significant discrepancy between clinical and neuropathology** 26 **diagnoses**

27 We observed a substantial discrepancy between the released clinical diagnoses and the
28 neuropathologically-confirmed diagnoses. Specifically, 300 (71%) of the 423 demented
29 participants had more than one pathology and 283 (67%) of them had been clinically diagnosed as

1 AD only. This observation is in line with recent clinical studies³¹ and postmortem autopsy
2 examinations^{30,32,94} reporting the frequent coexistence of AD with other neurodegenerative
3 disorders and highlighting the challenges in distinguishing the predominant pathology in clinical
4 dementia cases (AD versus VD^{75,95} or LBD⁹⁶).

5 **The deep learning model achieved the highest overall accuracy**

6 The best deep neural networks with asymmetric loss function achieved the overall best
7 performance compared to RF and ensemble SVM models, with a balanced accuracy of 0.844,
8 0.839, and 0.623 for AD, VD, and LBD, respectively. Additional parameters were a precision of
9 0.701, a recall of 0.793, and an F1 of 0.744. Although class imbalance is the most common
10 limitation of multi-label classification studies, in the present study, this issue was partially
11 mitigated by exploring the use of asymmetric loss during the training of our deep neural networks.
12 The balanced accuracy for LBD could be considered satisfactory, considering that structural
13 neuroimaging might exhibit a less characteristic signature for LBD pathology^{19,97}, which is more
14 associated with reduced functional signals in the occipital lobe and relative sparing of metabolism
15 in the posterior cingulate cortex in PET imaging^{19,97,98}.

16 As a comparison, the best ensemble SVM model achieved a lower balanced accuracy for AD and
17 VD, a much lower accuracy for LBD, a lower F1 score, and a lower averaged balanced accuracy
18 compared to the best deep learning model. The higher precision with a much lower recall observed
19 from the SVM model indicates that it is confident about easy positives but fails to detect hard
20 positives, like LBD, resulting in more false negatives. The reason for the relatively lower precision
21 of deep learning with an asymmetric loss function, but not with a binary cross-entropy loss
22 function, might stem from the probability shifting mechanism of the asymmetric loss function,
23 where it omits easy negative samples and thus might lead to more false positives. Overall, based
24 on the two composite metrics—F1 and $BACC_{avg}$ —the deep learning model achieved the best
25 performance.

26 As shown in **Supplementary Figure 1**, around half of the MRI scans used to train the model were
27 acquired in 1.5T scanners and the rest in 3T scanners. Including scanners of different field strengths
28 doubled our sample size and was intended to make our statistical models more robust to inter-
29 scanner differences. The influence of scanner properties and scanning protocol will be an essential

1 topic to explore while expanding our approach once data from diverse scanners are available in
2 the future.

3 The pathology probability charts derived from the machine learning model outputs established that
4 the model reaching the best classification performance was also the best at disentangling the three
5 neurodegenerative pathologies by achieving the highest silhouette score.

6 **Explainable deep-learning heatmaps revealed pathology-specific** 7 **patterns**

8 The brain alterations revealed by the group deep-learning heatmaps from the best deep neural
9 networks confirmed that the model was focusing on distinct brain regions when classifying the
10 three pathologies, and we found that these deep-learning heatmaps closely correspond to prior
11 findings. Specifically, the AD deep-learning heatmap accurately captured bilateral hippocampal
12 regions, consistent with numerous previous AD studies^{11,99,100}. Additionally, the VD deep-learning
13 heatmap highlighted many white matter regions commonly affected by WMH or other vascular
14 endophenotypes^{101–103}. Lastly, the LBD deep-learning heatmap focused on the occipital region.
15 This result is consistent with the fact that LBD is closely related to occipital changes^{20–22}. However,
16 previous studies investigating MRI-based structural atrophy in LBD have produced conflicting
17 results, with some studies only reporting atrophy in the limbic system^{104,105}. These contradictory
18 findings could potentially result from the presence of co-pathologies of AD.

19 Overall, the deep-learning heatmaps indicate that the deep neural networks could associate the
20 three neuropathology-confirmed diagnoses with distinct brain alterations, which was not the case
21 for the SVM model. The SVM coefficient maps captured more extensive changes throughout the
22 gray matter for all three diagnoses and failed to provide distinguishing information. This result is
23 consistent with our previous finding, demonstrating the advantage of capturing dementia patterns
24 using modern deep-learning heatmaps rather than the coefficients of SVM⁴⁵.

25 **DeepSPARE indices were associated with pathology-specific** 26 **biomarkers and measures**

27 Our deep learning-derived DeepSPARE indices presented diverse sensitivities to the three most
28 frequent co-pathologies. Significant associations with DeepSPARE-AD were observed for the
29 MMSE, TrailB, memory, hippocampal volume, the Braak stage, the CERAD score, and the Thal

1 phase. VD is associated with a higher prevalence of white matter diseases and vascular brain
2 injury^{101,106,107} and our results revealed significant associations of DeepSPARE-VD with CAA and
3 WMH. Lewy body stage was only significantly associated with DeepSPARE-LBD. These results
4 suggest that the DeepSPARE indices are sensitive and specific to well-established AD, VD, and
5 LBD biomarkers and measurements.

6 The out-of-sample validation in the ADNI dataset further strengthens the generalizability and
7 external validity of our findings. The AD/MCI_{Aβ+} group exhibited robust and consistent
8 associations with the tested cognitive, imaging, plasma, and CSF measures. These associations
9 were similar to those observed in the training/testing data. Specifically, DeepSPARE-AD was
10 associated with cognitive scores and hippocampal volume. In contrast, DeepSPARE-VD was
11 associated with WMH volume, confirming the ability of DeepSPARE to reflect changes related to
12 AD and VD pathology in an unseen dataset. Previous studies have demonstrated the ability of CSF
13 pTau181/Aβ42 ratio^{86,91} and pTau181^{90,108} to predict the early clinical decline and future AD
14 conversion. Our findings revealed consistent results where DeepSPARE-AD indices were
15 significantly associated with the CSF pTau181/Aβ42 ratio and pTau181. Additionally, recent
16 studies have shown that CSF α-synuclein SAA is a potential early biomarker for synucleinopathies
17 and can diagnose Parkinson's disease patients with high accuracy^{23,24,109,110}. Following these
18 reports, we observed a significant association of CSF α-synuclein SAA only with DeepSPARE-
19 LBD. Plasma biomarkers are particularly interesting because of their convenience and minimally
20 invasive nature. Plasma pTau181 has been proposed as a potential plasma biomarker for AD^{89,90}.
21 Correspondingly, we observed a significant association between DeepSPARE-AD and plasma
22 pTau181.

23 **Clinical relevance of the DeepSPARE indices**

24 Noninvasive pathology surrogate biomarkers could be crucial in advancing our understanding of
25 the heterogeneity and management of neurodegenerative disorders⁵⁰. Clinically, these indices
26 could serve as valuable imaging markers, particularly in cases of atypical, complex, and concurrent
27 dementia presentations. One striking finding of our current study is the improved understanding
28 of the role of WMH in dementia. Our previous research in the general population¹ demonstrated a
29 strong relationship between WMH and SPARE-AD values, a finding replicated in multiple
30 cohorts⁴⁶. However, since these training sets relied only on clinical diagnosis of AD and ignored

1 the possibilities of co-existing pathology, the previous indices were not able to distinguish
2 dementia of AD from VD^{1,46}. This disentanglement was only achievable when our advanced deep-
3 learning framework was developed, where we found a more specific role of WMH in VD as they
4 were associated only with DeepSPARE-VD when adjusted for other indices. Acknowledging that
5 these pathologies might co-exist, the DeepSPARE indices can provide sensitive and specific
6 markers of these pathologies.
7 Furthermore, from a clinical trial perspective, the DeepSPARE indices could help identify at-risk
8 populations with subclinical neuropathology, enabling diagnosis during the extended latency
9 period between neuropathological changes and the onset of clinical symptoms. Our previous
10 work¹¹¹ has simulated the effect of improving statistical power in clinical trials in prodromal
11 Alzheimer's disease. In future research, they may facilitate the identification of high-risk cohorts
12 and act as tools for monitoring neuropathology burden in developing and testing new or repurposed
13 disease-modifying treatments.

14 **Strengths and limitations**

15 This study builds on several strengths, such as the use of neuropathology-confirmed diagnoses
16 instead of clinical diagnoses, the adoption of multi-label deep learning networks to diagnose co-
17 pathologies based on antemortem 3D T1-weighted MRI scans with minimal pre-processing, and
18 the derivation of explainable deep-learning heatmaps to visualize the brain alteration patterns of
19 three most common co-pathologies from a single model. The combination of two large sets of
20 neuroimaging data to establish and validate the statistical models is also a strength of the present
21 study. Furthermore, the introduction of the new deep learning-derived DeepSPARE indices is
22 another notable strength. We have extensively validated DeepSPARE indices with a wide range of
23 clinical, imaging, and biofluid biomarkers and measures, demonstrating their abilities to reflect
24 AD, VD, and LBD pathologies individually and potentially serve as noninvasive pathology
25 biomarkers. Those indices, based on T1-weighted images, could exhibit broad applicability in both
26 clinical and research settings since T1 sequences are already part of the majority of clinical and
27 research MRI protocols.

28 The main limitation of the study resides in the data imbalance across neuropathology subgroups,
29 with a notably larger number of AD and VD patients compared to other subgroups. This disparity
30 could result in superior model detections for these two pathologies. Our current work demonstrated
31 the ability to differentiate AD, VD, and LBD with great sensitivity and precision from a single T1

1 MR scan, however, other pathologies, such as Limbic-predominant age-related TDP-43
2 encephalopathy (LATE)¹¹² and frontotemporal dementia, were underrepresented due to limited
3 sample sizes. In the meantime, we are actively working to expand our training sample and establish
4 more extensive databases that integrate neuroimaging with postmortem evaluations³⁰. Moreover,
5 the tau PET dataset was limited in the MCI_{Aβ}- and CN groups, making it challenging to compute
6 SPARE-Tau⁸⁸ indices for these populations. Additionally, integrating other imaging modalities
7 could have significantly improved the model accuracy since PET and FLAIR scans provide more
8 informative neuroimaging signatures for LBD¹⁹ and VD¹, respectively. Overall, additional data
9 will be required to reach a sufficient classification accuracy for use in routine clinical practice and
10 extend the models to other related dementias, such as LATE and frontotemporal dementias.

11 **Conclusion**

12 This study demonstrates the benefits of advanced deep learning models in detecting the three most
13 common dementia co-pathologies, revealing neuroimaging brain alteration patterns specific to
14 each pathology, and developing innovative deep learning-derived DeepSPARE indices. More
15 specifically, the deep learning model achieves the highest overall accuracy compared to other
16 machine learning models in identifying neuropathology-confirmed dementia diagnoses. The
17 explainable deep-learning heatmaps uncover distinctive structural neuroimaging brain alteration
18 patterns for AD in the hippocampus, VD in white matter, and LBD in occipital regions, which is
19 not the case for the traditional machine learning method. Lastly, the DeepSPARE indices
20 demonstrate effective, consistent, and robust pathology-specific associations across a variety of
21 cognitive tests, neuropathology, imaging, plasma, and CSF biomarkers and measures. They have
22 also been validated and replicated by an out-of-sample dataset. Clinically, the DeepSPARE indices
23 hold promise as noninvasive diagnostic tools, offering the ability to differentiate between co-
24 occurring neurodegenerative pathologies and identify at-risk populations before symptoms
25 become apparent. Furthermore, these indices could potentially enhance diagnostic accuracy
26 compared to the current clinical diagnostic criteria and thus provide treatment strategies for
27 specific pathologies.

28 **Data availability**

29 The NACC data supporting this study's findings are available from the National Alzheimer's
30 Coordinating Center. Restrictions apply to the availability of these data, which were used under

1 license for this study. NACC data are available at (<https://naccddata.org/>) with permission by
2 request. The ADNI data used in the present work were obtained from the Alzheimer's Disease
3 Neuroimaging Initiative database (<https://adni.loni.usc.edu/>). The code from this study is available
4 at <https://github.com/UTHSCSA-NAL/DeepSPARE>.

5 **Acknowledgments**

6 Part of the brain scans used in the present work were obtained from the Alzheimer's Disease
7 Neuroimaging Initiative (ADNI) database (<https://adni.loni.usc.edu/>), but ADNI investigators did
8 not participate in the design, the analysis, or the writing of the manuscript. A complete listing of
9 ADNI investigators can be found on their website (<https://adni.loni.usc.edu/>). J. T. is the Ann and
10 Billy Harrison Centennial Chair in Alzheimer's Research.

11 **Funding**

12 This study was supported in part by the National Institute of Health (NIH) grant P30AG066546
13 (South Texas Alzheimer's Disease Research Center) and grant numbers 5R01HL127659,
14 1U24AG074855, 5R01AG080821, 1R01AG085571, and 5R01AG083865. This study was
15 supported in part by the William and Ella Owens Medical Research Foundation.

16 The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by
17 the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar
18 Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas
19 Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD),
20 P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518
21 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott
22 Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD),
23 P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30
24 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI
25 Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric
26 Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD,
27 PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD,
28 FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD),
29 P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD,
30 MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30

1 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20
 2 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI
 3 Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James
 4 Leverenz, MD). The ADSP Phenotype Harmonization Consortium (ADSP-PHC) is funded by NIA
 5 (U24 AG074855, U01 AG068057 and R01 AG059716).

6 **Competing interests**

7 J.T. has served as a speaker for Eisai and a consultant for GE Healthcare. Other authors have
 8 declared that no competing interests exist.

9 **Supplementary material**

10 Supplementary material is available at *Brain* online.

11 **References**

- 12 1. Habes M, Erus G, Toledo JB, et al. White matter hyperintensities and imaging patterns of
 13 brain ageing in the general population. *Brain*. 2016;139(4):1164-1179.
- 14 2. Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA. Population
 15 estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United
 16 States (2020–2060). *Alzheimers Dement*. 2021;17(12):1966-1975. doi:10.1002/alz.12362
- 17 3. 2024 Alzheimer's disease facts and figures. *Alzheimers Dement*. Published online April 30,
 18 2024;alz.13809. doi:10.1002/alz.13809
- 19 4. Eschweiler GW, Leyhe T, Klöppel S, Hüll M. New developments in the diagnosis of
 20 dementia. *Dtsch Arztebl Int*. 2010;107(39):677.
- 21 5. Galvin JE, Sadowsky CH. Practical guidelines for the recognition and diagnosis of
 22 dementia. *J Am Board Fam Med*. 2012;25(3):367-382.
- 23 6. Robillard A. Clinical diagnosis of dementia. *Alzheimers Dement*. 2007;3(4):292-298.
- 24 7. Meyer JS, Huang J, Chowdhury MH. MRI confirms mild cognitive impairments prodromal
 25 for Alzheimer's, vascular and Parkinson-Lewy body dementias. *J Neurol Sci*. 2007;257(1-2):97-
 26 104.
- 27 8. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Initiative ADN. Machine learning
 28 framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *Neuroimage*.
 29 2015;104:398-412.

- 1 9. Bashyam VM, Erus G, Doshi J, et al. MRI signatures of brain age and disease over the
2 lifespan based on a deep brain network and 14 468 individuals worldwide. *Brain*.
3 2020;143(7):2312-2324.
- 4 10. Koikkalainen J, Rhodius-Meester H, Tolonen A, et al. Differential diagnosis of
5 neurodegenerative diseases using structural MRI data. *NeuroImage Clin*. 2016;11:435-449.
- 6 11. Dickerson BC, Goncharova I, Sullivan M, et al. MRI-derived entorhinal and hippocampal
7 atrophy in incipient and very mild Alzheimer's disease. *Neurobiol Aging*. 2001;22(5):747-754.
- 8 12. Imabayashi E, Matsuda H, Tabira T, et al. Comparison between brain CT and MRI for
9 voxel-based morphometry of Alzheimer's disease. *Brain Behav*. 2013;3(4):487-493.
10 doi:10.1002/brb3.146
- 11 13. Scheltens P, Leys D, Barkhof F, et al. Atrophy of medial temporal lobes on MRI in
12 "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological
13 correlates. *J Neurol Neurosurg Psychiatry*. 1992;55(10):967-972. doi:10.1136/jnnp.55.10.967
- 14 14. Habes M, Sotiras A, Erus G, et al. White matter lesions: Spatial heterogeneity, links to risk
15 factors, cognition, genetics, and atrophy. *Neurology*. 2018;91(10).
16 doi:10.1212/WNL.0000000000006116
- 17 15. Habes M, Erus G, Toledo JB, et al. Regional tract-specific white matter hyperintensities
18 are associated with patterns of aging-related brain atrophy via vascular risk factors, but also
19 independently. Wolk D, Villemagne V, Dickerson B, eds. *Alzheimers Dement Diagn Assess Dis*
20 *Monit*. 2018;10(1):278-284. doi:10.1016/j.dadm.2018.02.002
- 21 16. Chu B, Kampschulte A, Ferguson MS, et al. Hemorrhage in the Atherosclerotic Carotid
22 Plaque: A High-Resolution MRI Study. *Stroke*. 2004;35(5):1079-1084.
23 doi:10.1161/01.STR.0000125856.25309.86
- 24 17. Chauhan G, Adams HHH, Satizabal CL, et al. Genetic and lifestyle risk factors for MRI-
25 defined brain infarcts in a population-based setting. *Neurology*. 2019;92(5).
26 doi:10.1212/WNL.0000000000006851
- 27 18. Charisis S, Rashid T, Liu H, et al. Assessment of Risk Factors and Clinical Importance of
28 Enlarged Perivascular Spaces by Whole-Brain Investigation in the Multi-Ethnic Study of
29 Atherosclerosis. *JAMA Netw Open*. 2023;6(4):e239196.
30 doi:10.1001/jamanetworkopen.2023.9196

- 1 19. McKeith IG, Boeve BF, Dickson DW, et al. Diagnosis and management of dementia with
2 Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology*. 2017;89(1):88-100.
- 3 20. Burton EJ, Karas G, Paling S, et al. Patterns of cerebral atrophy in dementia with Lewy
4 bodies using voxel-based morphometry. *Neuroimage*. 2002;17(2):618-630.
- 5 21. Borroni B, Anchisi D, Paghera B, et al. Combined 99mTc-ECD SPECT and
6 neuropsychological studies in MCI for the assessment of conversion to AD. *Neurobiol Aging*.
7 2006;27(1):24-31. doi:10.1016/j.neurobiolaging.2004.12.010
- 8 22. Watson R, O'Brien JT, Barber R, Blamire AM. Patterns of gray matter atrophy in dementia
9 with Lewy bodies: a voxel-based morphometry study. *Int Psychogeriatr*. 2012;24(4):532-540.
- 10 23. Bräuer S, Rossi M, Sajapin J, et al. Kinetic parameters of alpha-synuclein seed
11 amplification assay correlate with cognitive impairment in patients with Lewy body disorders.
12 *Acta Neuropathol Commun*. 2023;11(1):162. doi:10.1186/s40478-023-01653-3
- 13 24. Yoo D, Bang JI, Ahn C, et al. Diagnostic value of α -synuclein seeding amplification assays
14 in α -synucleinopathies: A systematic review and meta-analysis. *Parkinsonism Relat Disord*.
15 2022;104:99-109. doi:10.1016/j.parkreldis.2022.10.007
- 16 25. Atri A. The Alzheimer's disease clinical spectrum: diagnosis and management. *Med Clin*.
17 2019;103(2):263-293.
- 18 26. Chong MS, Sahadevan S. Preclinical Alzheimer's disease: diagnosis and prediction of
19 progression. *Lancet Neurol*. 2005;4(9):576-579.
- 20 27. Binnewijzend MA, Kuijer JP, van der Flier WM, et al. Distinct perfusion patterns in
21 Alzheimer's disease, frontotemporal dementia and dementia with Lewy bodies. *Eur Radiol*.
22 2014;24:2326-2333.
- 23 28. Filippi M, Agosta F, Barkhof F, et al. EFNS task force: the use of neuroimaging in the
24 diagnosis of dementia. *Eur J Neurol*. 2012;19(12):1487-1501.
- 25 29. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-
26 based classification studies and associated feature extraction methods for Alzheimer's disease and
27 its prodromal stages. *NeuroImage*. 2017;155:530-548.
- 28 30. Li K, Rashid T, Li J, et al. Postmortem Brain Imaging in Alzheimer's Disease and Related
29 Dementias: The South Texas Alzheimer's Disease Research Center Repository. *J Alzheimers Dis*
30 *JAD*. 2023;96(3):1267-1283. doi:10.3233/JAD-230389

- 1 31. Knopman DS, Parisi JE, Boeve BF, et al. Vascular dementia in a population-based autopsy
2 study. *Arch Neurol*. 2003;60(4):569-575.
- 3 32. Jellinger KA, Attems J. Prevalence of dementia disorders in the oldest-old: an autopsy
4 study. *Acta Neuropathol (Berl)*. 2010;119:421-433.
- 5 33. Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of
6 Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J*
7 *Neuropathol Exp Neurol*. 2012;71(4):266-273.
- 8 34. Rizzo G, Arcuti S, Copetti M, et al. Accuracy of clinical diagnosis of dementia with Lewy
9 bodies: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry*. 2018;89(4):358-
10 366.
- 11 35. Skogseth R, Hortobágyi T, Soennesyn H, et al. Accuracy of clinical diagnosis of dementia
12 with Lewy bodies versus neuropathology. *J Alzheimers Dis*. 2017;59(4):1139-1152.
- 13 36. Toledo JB, Brettschneider J, Grossman M, et al. CSF biomarkers cutoffs: the importance
14 of coincident neuropathological diseases. *Acta Neuropathol (Berl)*. 2012;124(1):23-35.
15 doi:10.1007/s00401-012-0983-7
- 16 37. Mueller SG, Weiner MW, Thal LJ, et al. The Alzheimer's disease neuroimaging initiative.
17 *Neuroimaging Clin*. 2005;15(4):869-877.
- 18 38. Suk HI, Shen D. Deep learning-based feature representation for AD/MCI classification. In:
19 *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013: 16th*
20 *International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II 16*.
21 Springer; 2013:583-590.
- 22 39. Venugopalan J, Tong L, Hassanzadeh HR, Wang MD. Multimodal deep learning models
23 for early detection of Alzheimer's disease stage. *Sci Rep*. 2021;11(1):3254.
- 24 40. Pellegrini E, Ballerini L, Hernandez M del CV, et al. Machine learning of neuroimaging
25 for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers*
26 *Dement Diagn Assess Dis Monit*. 2018;10:519-535.
- 27 41. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng*.
28 2017;19(1):221-248. doi:10.1146/annurev-bioeng-071516-044442
- 29 42. Aggarwal R, Sounderajah V, Martin G, et al. Diagnostic accuracy of deep learning in
30 medical imaging: a systematic review and meta-analysis. *Npj Digit Med*. 2021;4(1):65.
31 doi:10.1038/s41746-021-00438-z

- 1 43. Dyrba M, Hanzig M, Altenstein S, et al. Improving 3D convolutional neural network
2 comprehensibility via interactive visualization of relevance maps: evaluation in Alzheimer's
3 disease. *Alzheimers Res Ther.* 2021;13(1):191. doi:10.1186/s13195-021-00924-2
- 4 44. Böhle M, Eitel F, Weygandt M, Ritter K. Layer-wise relevance propagation for explaining
5 deep neural network decisions in MRI-based Alzheimer's disease classification. *Front Aging*
6 *Neurosci.* 2019;11:194.
- 7 45. Wang D, Honnorat N, Fox PT, et al. Deep neural network heatmaps capture Alzheimer's
8 disease patterns reported in a large meta-analysis of neuroimaging studies. *Neuroimage.*
9 2023;269:119929.
- 10 46. Habes M, Pomponio R, Shou H, et al. The Brain Chart of Aging: Machine-learning
11 analytics reveals links between brain aging, white matter disease, amyloid burden, and cognition
12 in the iSTAGING consortium of 10,216 harmonized MR scans. *Alzheimers Dement.*
13 2021;17(1):89-102. doi:10.1002/alz.12178
- 14 47. Davatzikos C, Xu F, An Y, Fan Y, Resnick SM. Longitudinal progression of Alzheimer's-
15 like patterns of atrophy in normal older adults: the SPARE-AD index. *Brain J Neurol.* 2009;132(Pt
16 8):2026-2035. doi:10.1093/brain/awp091
- 17 48. Habes M, Janowitz D, Erus G, et al. Advanced brain aging: relationship with epidemiologic
18 and genetic risk factors, and overlap with Alzheimer disease atrophy patterns. *Transl Psychiatry.*
19 2016;6:e775. doi:10.1038/tp.2016.39
- 20 49. Bashyam VM, Doshi J, Erus G, et al. Deep Generative Medical Image Harmonization for
21 Improving Cross-Site Generalization in Deep Learning Predictors. *J Magn Reson Imaging.*
22 2022;55(3):908-916. doi:10.1002/jmri.27908
- 23 50. Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C. Disentangling
24 heterogeneity in Alzheimer's disease and related dementias using data-driven methods. *Biol*
25 *Psychiatry.* 2020;88(1):70-82.
- 26 51. Beekly DL, Ramos EM, Lee WW, et al. The National Alzheimer's Coordinating Center
27 (NACC) database: the uniform data set. *Alzheimer Dis Assoc Disord.* 2007;21(3):249-258.
- 28 52. Ho TK. Random decision forests. In: *Proceedings of 3rd International Conference on*
29 *Document Analysis and Recognition.* Vol 1. IEEE; 1995:278-282.
- 30 53. Smola AJ, Schölkopf B. A Tutorial on Support Vector Regression. *Stat Comput Arch.*
31 2004;14(3):199-222.

- 1 54. Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta*
2 *Neuropathol (Berl)*. 1991;82(4):239-259. doi:10.1007/BF00308809
- 3 55. Thal DR, Rüb U, Orantes M, Braak H. Phases of A β -deposition in the human brain and its
4 relevance for the development of AD. *Neurology*. 2002;58(12):1791-1800.
- 5 56. DeTure MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. *Mol*
6 *Neurodegener*. 2019;14(1):1-18.
- 7 57. Montine TJ, Phelps CH, Beach TG, et al. National Institute on Aging-Alzheimer's
8 Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical
9 approach. *Acta Neuropathol (Berl)*. 2012;123(1):1-11. doi:10.1007/s00401-011-0910-3
- 10 58. T O'Brien J, Thomas A. Vascular dementia. *The Lancet*. 2015;386(10004):1698-1706.
- 11 59. Doshi J, Erus G, Ou Y, et al. MUSE: MUlti-atlas region Segmentation utilizing Ensembles
12 of registration algorithms and parameters, and locally optimal atlas selection. *NeuroImage*.
13 2016;127:186-195.
- 14 60. Fonov V, Evans A, McKinstry R, Almlí C, Collins D. Unbiased nonlinear average age-
15 appropriate brain templates from birth to adulthood. *NeuroImage*. 2009;47:S102.
- 16 61. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image
17 registration with cross-correlation: evaluating automated labeling of elderly and
18 neurodegenerative brain. *Med Image Anal*. 2008;12(1):26-41.
- 19 62. Avants BB, Tustison NJ, Wu J, Cook PA, Gee JC. An open source multivariate framework
20 for n-tissue segmentation with evaluation on public data. *Neuroinformatics*. 2011;9(4):381-400.
- 21 63. Sabottke CF, Spieler BM. The Effect of Image Resolution on Deep Learning in
22 Radiography. *Radiol Artif Intell*. 2020;2(1):e190015. doi:10.1148/ryai.2019190015
- 23 64. Touvron H, Vedaldi A, Douze M, Jégou H. Fixing the train-test resolution discrepancy.
24 Published online 2019. doi:10.48550/ARXIV.1906.06423
- 25 65. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In:
26 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE;
27 2016:770-778.
- 28 66. Ben-Baruch E, Ridnik T, Zamir N, et al. Asymmetric Loss For Multi-Label Classification.
29 Published online 2021.

- 1 67. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more
2 correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*.
3 1988;44(3):837-845.
- 4 68. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster
5 analysis. *J Comput Appl Math*. 1987;20:53-65. doi:10.1016/0377-0427(87)90125-7
- 6 69. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J*
7 *Mach Learn Res*. 2011;12:2825-2830.
- 8 70. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks. In: *International*
9 *Conference on Machine Learning*. IEEE; 2017:3319-3328.
- 10 71. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual
11 Explanations From Deep Networks via Gradient-Based Localization. In: *Proceedings of the IEEE*
12 *International Conference on Computer Vision*. ; 2017:618-626.
- 13 72. Folstein MF, Robins LN, Helzer JE. The mini-mental state examination. *Arch Gen*
14 *Psychiatry*. 1983;40(7):812-812.
- 15 73. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education.
16 *Arch Clin Neuropsychol*. 2004;19(2):203-214.
- 17 74. Wechsler D. Wechsler memory scale-revised. *Psychol Corp*. Published online 1987.
- 18 75. Gorelick PB, Scuteri A, Black SE, et al. Vascular contributions to cognitive impairment
19 and dementia: a statement for healthcare professionals from the American Heart
20 Association/American Stroke Association. *stroke*. 2011;42(9):2672-2713.
- 21 76. Biffi A, Greenberg SM. Cerebral Amyloid Angiopathy: A Systematic Review. *J Clin*
22 *Neurol*. 2011;7(1):1. doi:10.3988/jcn.2011.7.1.1
- 23 77. Fillenbaum GG, van Belle G, Morris JC, et al. Consortium to Establish a Registry for
24 Alzheimer's Disease (CERAD): the first twenty years. *Alzheimers Dement J Alzheimers Assoc*.
25 2008;4(2):96-109. doi:10.1016/j.jalz.2007.08.005
- 26 78. Marui W, Iseki E, Nakai T, et al. Progression and staging of Lewy pathology in brains from
27 patients with dementia with Lewy bodies. *J Neurol Sci*. 2002;195(2):153-159. doi:10.1016/S0022-
28 510X(02)00006-0
- 29 79. Rajapakse JC, Giedd JN, DeCarli C, et al. A technique for single-channel MR brain tissue
30 segmentation: Application to a pediatric sample. *Magn Reson Imaging*. 1996;14(9):1053-1065.
31 doi:10.1016/S0730-725X(96)00113-0

- 1 80. Fletcher E, Singh B, Harvey D, Carmichael O, DeCarli C. Adaptive image segmentation
2 for robust measurement of longitudinal brain tissue change. In: *2012 Annual International*
3 *Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE; 2012:5319-5322.
4 doi:10.1109/EMBC.2012.6347195
- 5 81. DeCarli C, Miller BL, Swan GE, et al. Predictors of Brain Morphology for the Men of the
6 NHLBI Twin Study. *Stroke*. 1999;30(3):529-536. doi:10.1161/01.STR.30.3.529
- 7 82. Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. McGraw-
8 hill; 2005.
- 9 83. Sheather S. *A Modern Approach to Regression with R*. Springer New York; 2009.
10 doi:10.1007/978-0-387-09608-7
- 11 84. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful
12 Approach to Multiple Testing. *J R Stat Soc Ser B Stat Methodol*. 1995;57(1):289-300.
13 doi:10.1111/j.2517-6161.1995.tb02031.x
- 14 85. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using
15 **lme4**. *J Stat Softw*. 2015;67(1). doi:10.18637/jss.v067.i01
- 16 86. Blennow K, Shaw LM, Stomrud E, et al. Predicting clinical decline and conversion to
17 Alzheimer's disease or dementia using novel Elecsys A β (1–42), pTau and tTau CSF
18 immunoassays. *Sci Rep*. 2019;9(1):19024. doi:10.1038/s41598-019-54204-z
- 19 87. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's Disease Assessment
20 Scale–Cognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia
21 Populations. A Narrative Review. *J Alzheimers Dis*. 2018;63(2):423-444. doi:10.3233/JAD-
22 170991
- 23 88. Toledo JB, Rashid T, Liu H, et al. SPARE-Tau: A flortaucipir machine-learning derived
24 early predictor of cognitive decline. *PloS One*. 2022;17(11):e0276392.
25 doi:10.1371/journal.pone.0276392
- 26 89. Olsson B, Lautner R, Andreasson U, et al. CSF and blood biomarkers for the diagnosis of
27 Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol*. 2016;15(7):673-684.
28 doi:10.1016/S1474-4422(16)00070-3
- 29 90. Shen XN, Huang YY, Chen SD, et al. Plasma phosphorylated-tau181 as a predictive
30 biomarker for Alzheimer's amyloid, tau and FDG PET status. *Transl Psychiatry*. 2021;11(1):585.
31 doi:10.1038/s41398-021-01709-9

- 1 91. Fagan AM, Roe CM, Xiong C, Mintun MA, Morris JC, Holtzman DM. Cerebrospinal Fluid
2 tau/ β -Amyloid42 Ratio as a Prediction of Cognitive Decline in Nondemented Older Adults. *Arch*
3 *Neurol.* 2007;64(3):343. doi:10.1001/archneur.64.3.noc60123
- 4 92. Palmqvist S, Rossi M, Hall S, et al. Cognitive effects of Lewy body pathology in clinically
5 unimpaired individuals. *Nat Med.* 2023;29(8):1971-1978. doi:10.1038/s41591-023-02450-0
- 6 93. Siderowf A, Concha-Marambio L, Lafontant DE, et al. Assessment of heterogeneity among
7 participants in the Parkinson's Progression Markers Initiative cohort using α -synuclein seed
8 amplification: a cross-sectional study. *Lancet Neurol.* 2023;22(5):407-417. doi:10.1016/S1474-
9 4422(23)00109-6
- 10 94. Brenowitz WD, Hubbard RA, Keene CD, et al. Mixed neuropathologies and estimated
11 rates of clinical progression in a large autopsy sample. *Alzheimers Dement.* 2017;13(6):654-662.
- 12 95. Sachdev P, Kalaria R, O'Brien J, et al. Diagnostic criteria for vascular cognitive disorders:
13 a VASCOG statement. *Alzheimer Dis Assoc Disord.* 2014;28(3):206-218.
- 14 96. Tiraboschi P, Salmon DP, Hansen LA, Hofstetter RC, Thal LJ, Corey-Bloom J. What best
15 differentiates Lewy body from Alzheimer's disease in early-stage dementia? *Brain.*
16 2006;129(3):729-735.
- 17 97. Walker Z, Possin KL, Boeve BF, Aarsland D. Lewy body dementias. *The Lancet.*
18 2015;386(10004):1683-1697.
- 19 98. Graff-Radford J, Murray ME, Lowe VJ, et al. Dementia with Lewy bodies: basis of
20 cingulate island sign. *Neurology.* 2014;83(9):801-809. doi:10.1212/WNL.0000000000000734
- 21 99. Barnes LL, Wilson RS, Bienias JL, Schneider JA, Evans DA, Bennett DA. Sex differences
22 in the clinical manifestations of Alzheimer disease pathology. *Arch Gen Psychiatry.*
23 2005;62(6):685-691. doi:10.1001/archpsyc.62.6.685
- 24 100. Pini L, Pievani M, Bocchetta M, et al. Brain atrophy in Alzheimer's Disease and aging.
25 *Ageing Res Rev.* 2016;30:25-48. doi:10.1016/j.arr.2016.01.002
- 26 101. Barber R, Scheltens P, Gholkar A, et al. White matter lesions on magnetic resonance
27 imaging in dementia with Lewy bodies, Alzheimer's disease, vascular dementia, and normal aging.
28 *J Neurol Neurosurg Psychiatry.* 1999;67(1):66-72.
- 29 102. Englund E. Neuropathology of white matter changes in Alzheimer's disease and vascular
30 dementia. *Dement Geriatr Cogn Disord.* 1998;9(Suppl. 1):6-12.

- 1 103. Gootjes L, Teipel S, Zebuhr Y, et al. Regional distribution of white matter hyperintensities
2 in vascular dementia, Alzheimer's disease and healthy aging. *Dement Geriatr Cogn Disord*.
3 2004;18(2):180-188.
- 4 104. Whitwell JL, Jack CR, Przybelski SA, et al. Temporoparietal atrophy: A marker of AD
5 pathology independent of clinical diagnosis. *Neurobiol Aging*. 2011;32(9):1531-1541.
6 doi:10.1016/j.neurobiolaging.2009.10.012
- 7 105. Nedelska Z, Ferman TJ, Boeve BF, et al. Pattern of brain atrophy rates in autopsy-
8 confirmed dementia with Lewy bodies. *Neurobiol Aging*. 2015;36(1):452-461.
- 9 106. Zarei M, Damoiseaux JS, Morgese C, et al. Regional white matter integrity differentiates
10 between vascular dementia and Alzheimer disease. *Stroke*. 2009;40(3):773-779.
- 11 107. Palesi F, De Rinaldis A, Vitali P, et al. Specific patterns of white matter alterations help
12 distinguishing Alzheimer's and vascular dementia. *Front Neurosci*. 2018;12:274.
- 13 108. Mielke MM, Hagen CE, Xu J, et al. Plasma phospho-tau181 increases with Alzheimer's
14 disease clinical severity and is associated with tau- and amyloid-positron emission tomography.
15 *Alzheimers Dement*. 2018;14(8):989-997. doi:10.1016/j.jalz.2018.02.013
- 16 109. Fernandes Gomes B, Farris CM, Ma Y, et al. α -Synuclein seed amplification assay as a
17 diagnostic tool for parkinsonian disorders. *Parkinsonism Relat Disord*. 2023;117:105807.
18 doi:10.1016/j.parkreldis.2023.105807
- 19 110. Bellomo G, De Luca CMG, Paoletti FP, Gaetani L, Moda F, Parnetti L. α -Synuclein Seed
20 Amplification Assays for Diagnosing Synucleinopathies: The Way Forward. *Neurology*.
21 2022;99(5):195-205. doi:10.1212/WNL.0000000000200878
- 22 111. Lou C, Habes M, Illenberger NA, et al. Leveraging machine learning predictive biomarkers
23 to augment the statistical power of clinical trials with baseline magnetic resonance imaging. *Brain*
24 *Commun*. 2021;3(4):fcab264. doi:10.1093/braincomms/fcab264
- 25 112. Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43
26 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503-1527.
27 doi:10.1093/brain/awz099

28

1 **Figure legends**

2 **Figure 1 Overlap of pathology and the discrepancy between neuropathology diagnosis and**
 3 **clinical diagnosis in demented participants.** (A) Overlap between the three pathologies in the
 4 423 neuropathology-confirmed participants with dementia (Alzheimer's disease (AD), vascular
 5 dementia (VD), and Lewy body dementia (LBD)). Those 423 demented and 361 control
 6 participants (CN) were used for machine learning model training. (B) For the 423 neuropathology-
 7 confirmed participants, the comparison between the neuropathology-confirmed diagnoses and the
 8 clinical diagnoses released by the ADNI/NACC consortia. 69 participants had received a clinical
 9 diagnosis of mild cognitive impairment (MCI), and 14 participants were diagnosed with other
 10 brain diseases (others).

11
 12 **Figure 2 ROC curves and pathology probability charts for three tested models.** (A) ROC
 13 curves and their AUC of AD, VD, and LBD, for the best RF, the best ensemble linear SVM, and
 14 the best deep neural networks. (B) Pathology probability charts of the test dataset obtained for the
 15 best RF, ensemble SVM, and deep neural networks. The first column of those probability charts
 16 estimates the ability of the model to distinguish AD and VD. The second column reflects the
 17 separation between AD and LBD. The last column indicates if the models are able to disentangle
 18 VD and LBD. The corresponding silhouette scores are also reported.

19
 20 **Figure 3 Explainable deep-learning heatmaps and SVM coefficient maps for the AD, VD,**
 21 **and LBD pathology.** Explainable deep-learning heatmaps reflect the patterns of brain alterations
 22 extracted from the best deep neural networks using the integrated gradients method. The SVM
 23 coefficient maps provide the patterns captured by the best linear SVM model. All values are
 24 normalized between 0 and 1.

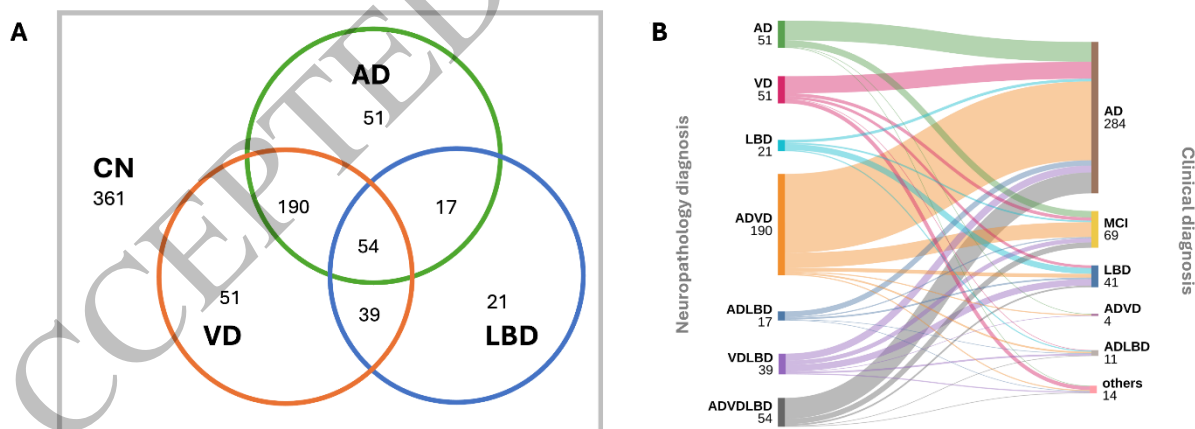
25
 26 **Figure 4 For the model test dataset, results from linear mixed models for the associations**
 27 **between cognitive, imaging, and neuropathology-based measures and DeepSPARE indices.**
 28 The number of available samples n is reported in each plot as well as the effect estimates (β
 29 coefficients) with 95% confidence intervals, FDR-corrected p -values (p_{FDR}), and variance
 30 inflation factors (VIF). Bold p -values indicate statistical significance ($p_{FDR} < 0.05$). LOGIMEM

1 = immediate logical memory subtest of the Wechsler Memory Scale-Revised; WMH = White
2 matter hyperintensity.

3
4 **Figure 5** For the out-of-sample ADNI validation dataset, results from linear mixed models
5 for the associations between cognitive, and imaging-based measures and DeepSPARE
6 indices. The number of available samples n is reported in each plot, as well as the effect estimates
7 (β coefficients) with 95% confidence intervals, FDR-corrected p -values (p_{FDR}), and variance
8 inflation factors (VIF). Bold p -values indicate statistical significance ($p_{FDR} < 0.05$). WMH =
9 White matter hyperintensity.

10
11 **Figure 6** For the out-of-sample ADNI validation dataset, results from linear mixed models
12 for the associations between biofluid measures and DeepSPARE indices. The number of
13 available samples n is reported in each plot, as well as the effect estimates (log odds ratios for α -
14 synuclein and β coefficients for other measures) with 95% confidence intervals, FDR-corrected
15 p -values (p_{FDR}), and variance inflation factors (VIF). Bold p -values indicate statistical significance
16 ($p_{FDR} < 0.05$).

17



18
19
20
21
Figure 1
159x58 mm (x DPI)

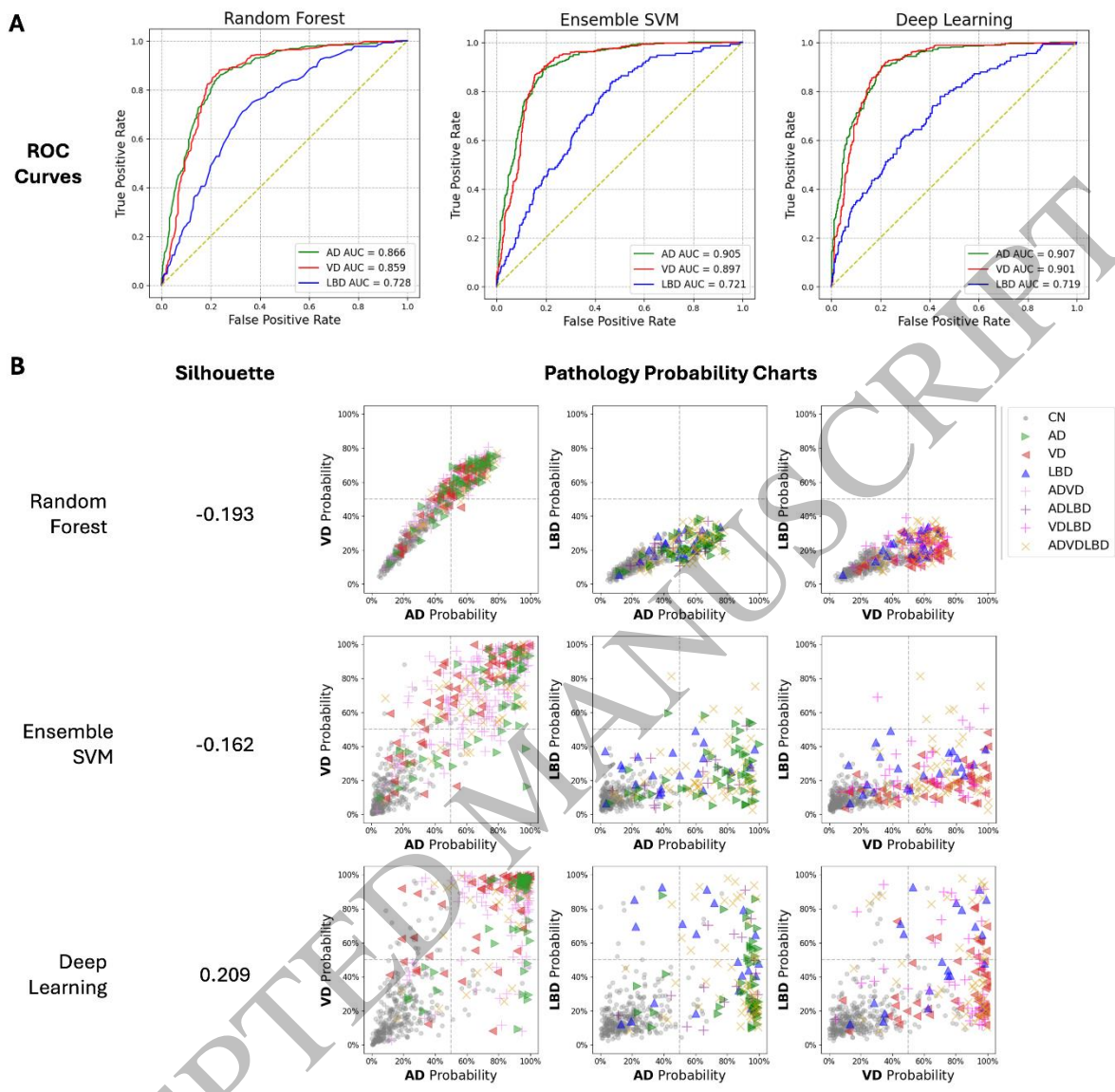


Figure 2
159x152 mm (x DPI)

1
2
3

1
2
3

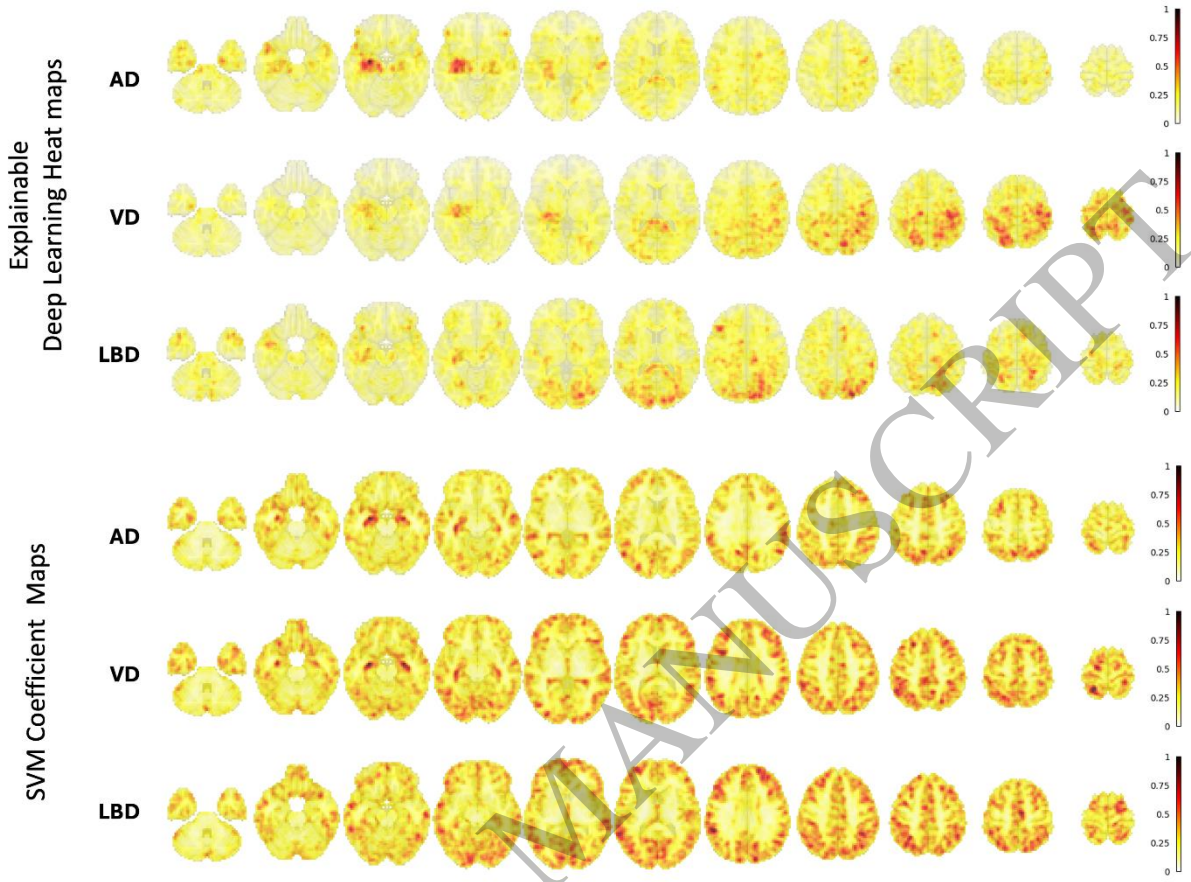


Figure 3
159x120 mm (x DPI)

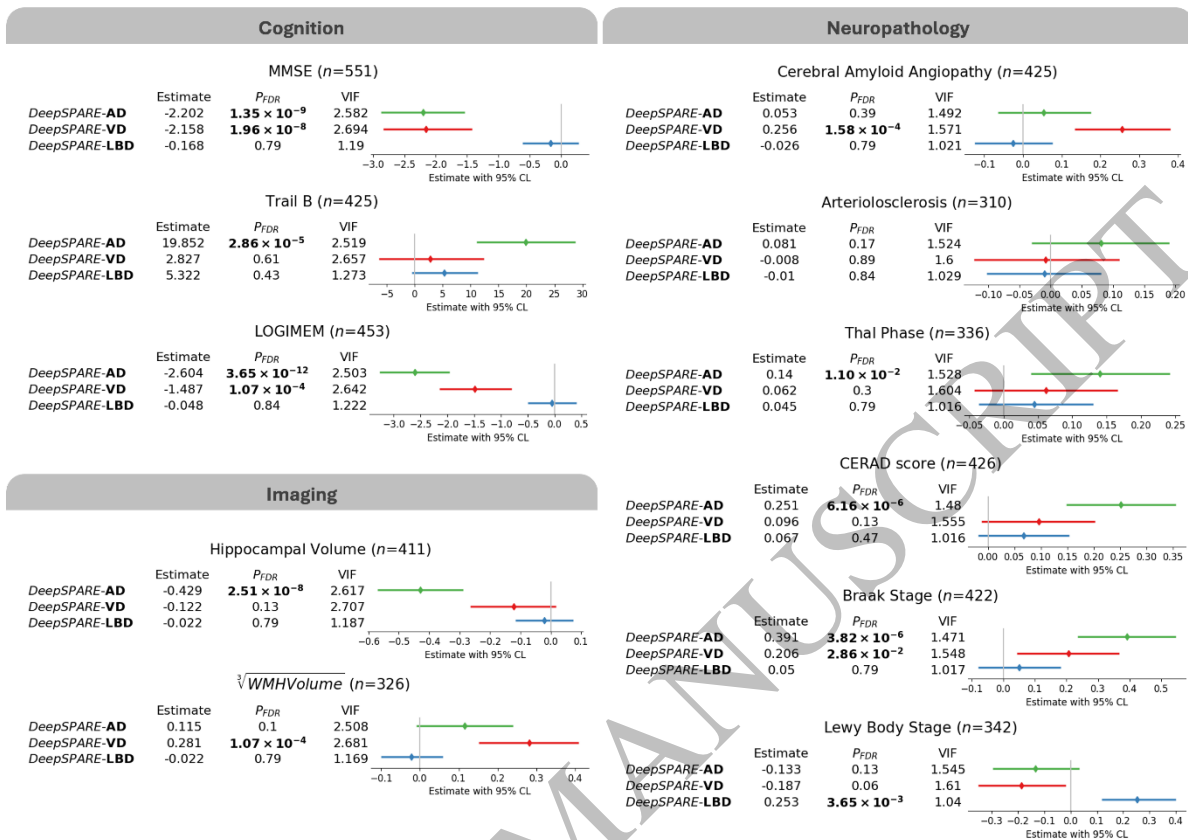


Figure 4
159x112 mm (x DPI)

1
2
3

Cognition
Imaging

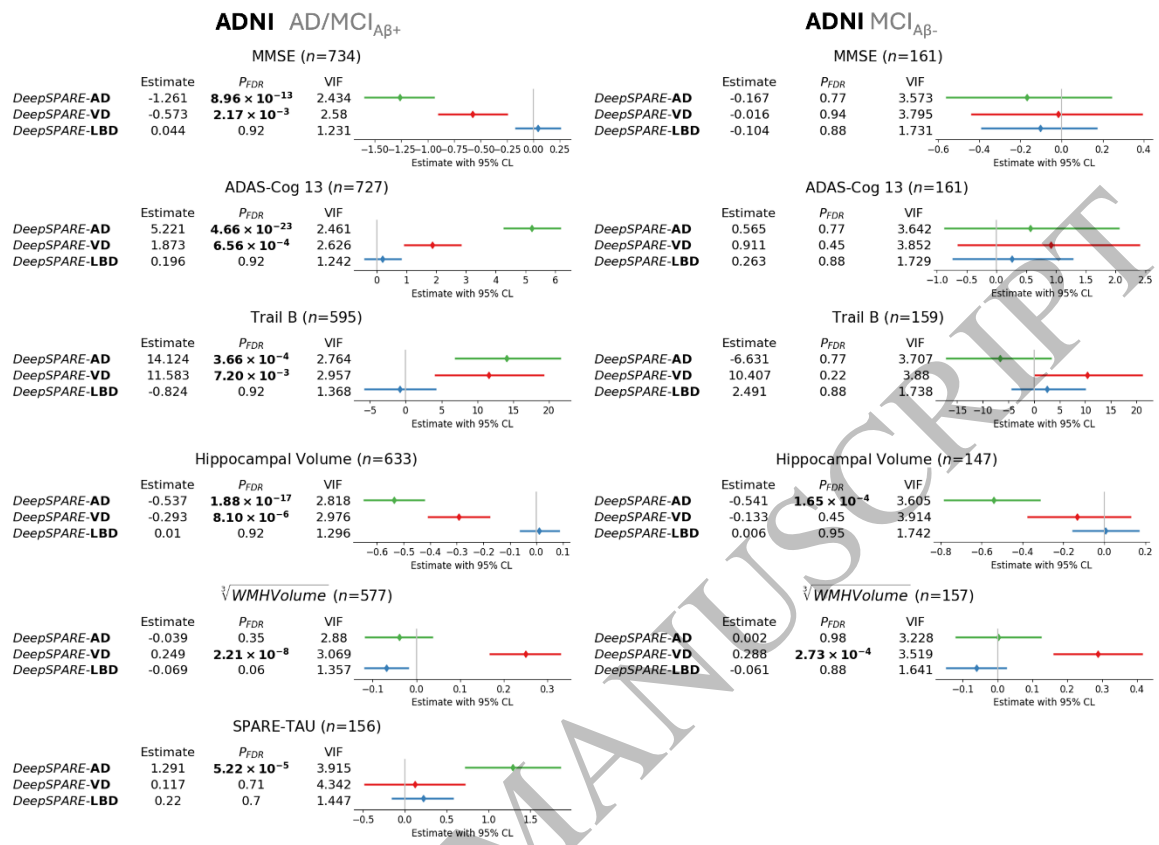


Figure 5
159x113 mm (x DPI)

1
2
3

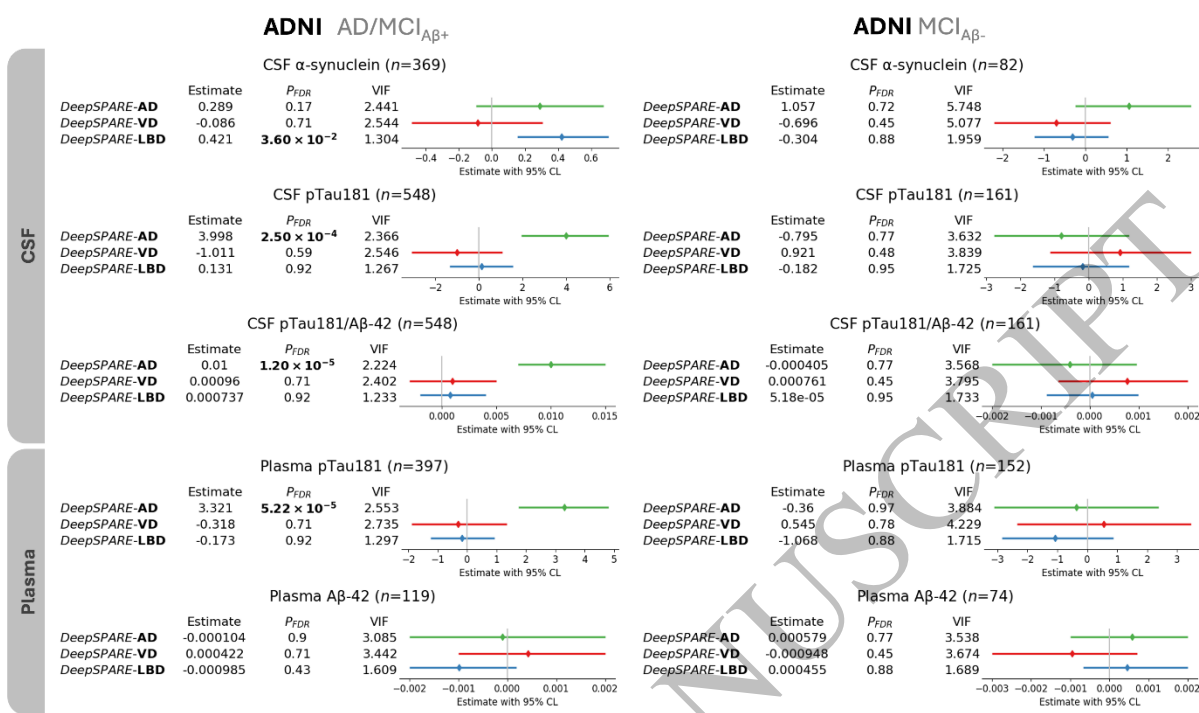


Figure 6
159x95 mm (x DPI)

Table 1 Participant characteristics by neuropathological diagnosis in the training and validation datasets

Participants	Number	Mean Age (std)	Sex (F/M)	Mean Education (std)	Mean MMSE (std)
Model Training dataset					
AD pathology	312	75.4 (8.9)	123/189	15.6 (3.0)	20.6 (6.1)
NACC	248	74.5 (9.2)	105/143	15.5 (3.1)	20.8 (6.2)
ADNI	64	78.8 (6.8)	18/46	16.7 (2.4)	19.4 (5.5)
VD pathology	334	76.6 (9.0)	121/213	15.4 (3.3)	21.15 (6.0)
NACC	278	75.9 (9.3)	108/170	15.3 (3.3)	21.2 (6.1)
ADNI	56	79.9 (7.0)	13/43	16.8 (2.6)	20.5 (5.4)
LBD pathology	131	75.8 (8.8)	33/98	15.9 (2.9)	22.2 (6.0)
NACC	101	74.8 (8.9)	30/71	15.6 (2.8)	22.4 (6.1)
ADNI	30	79.3 (7.3)	3/27	18.2 (1.7)	20.5 (4.7)
Control	361	75.5 (8.9)	130/231	15.8 (3.5)	28.6 (1.6)
NACC	291	74.9 (9.5)	111/180	15.7 (3.7)	28.5 (1.7)
ADNI	70	77.9 (5.3)	19/51	16.9 (2.7)	29.0 (1.2)
Out-of-sample Validation dataset					
Total	1041	72.1 (7.1)	460/581	16.2 (2.6)	26.3 (3.1)
AD/MCI _{Aβ+}	734	72.8 (8.5)	306/428	15.8 (2.7)	25.1 (3.3)
MCI _{Aβ-}	161	70.4 (7.5)	83/78	16.3 (2.6)	28.5 (1.5)
Control	146	71.4 (6.1)	71/75	16.6 (2.5)	29.1 (1.1)

AD = Alzheimer's disease; VD = vascular dementia; LBD = Lewy body dementia; MCI_{Aβ+} = mild cognitive impairment amyloid-β positive; MCI_{Aβ-} = mild cognitive impairment amyloid-β negative.

1
2
3
4
5
6
7
8

1
2
3
4
5
Table 2 Ten-fold cross-validated precision, recall, F1, balanced accuracies, and AUC of all the pathology classification models tested in this work

Model	Loss	Parameter	Precision	Recall	F1	BAC _{AD}	BACC _{VD}	BACC _{BD}	BACC _{avg}	AUC _{AD}	AUC _{VD}	AUC _{BD}
ResNet18	AL	b=16	0.722	0.736	0.729	0.820	0.820	0.606	0.749	0.905	0.883	0.702
		b=8 ^{a,b}	0.701	0.793	0.744	0.844	0.839	0.623	0.769	0.907	0.901	0.719
		b=4	0.699	0.743	0.720	0.824	0.818	0.594	0.745	0.896	0.893	0.719
	BCE	b=16	0.760	0.665	0.710	0.816	0.799	0.566	0.727	0.897	0.892	0.746
		b=8	0.766	0.676	0.718	0.823	0.817	0.547	0.729	0.906	0.893	0.732
		b=4	0.746	0.717	0.731	0.828	0.817	0.603	0.750	0.907	0.886	0.740
Ensemble Linear SVM	Hinge loss	c=10	0.783	0.673	0.724	0.822	0.832	0.531	0.728	0.904	0.897	0.714
		c=1	0.783	0.673	0.724	0.822	0.832	0.531	0.728	0.905	0.897	0.715
		c=0.1	0.783	0.673	0.724	0.822	0.832	0.531	0.728	0.905	0.897	0.713
		c=0.01 ^a	0.783	0.673	0.724	0.822	0.832	0.531	0.728	0.905	0.897	0.721
		c=0.001	0.795	0.658	0.720	0.810	0.836	0.504	0.717	0.883	0.876	0.737
		c=0.0001	0.758	0.367	0.494	0.636	0.702	0.500	0.613	0.843	0.819	0.735
RF	Gini impurity	n=20	0.730	0.519	0.606	0.724	0.741	0.499	0.655	0.839	0.818	0.680
		n=50	0.747	0.601	0.666	0.755	0.795	0.500	0.683	0.858	0.846	0.701
		n=100	0.759	0.611	0.677	0.786	0.784	0.500	0.690	0.861	0.848	0.710
		n=200	0.752	0.624	0.682	0.779	0.797	0.500	0.692	0.868	0.858	0.739
		n=300 ^a	0.764	0.625	0.688	0.786	0.802	0.500	0.696	0.866	0.859	0.728
		n=400	0.758	0.629	0.688	0.785	0.802	0.500	0.696	0.867	0.854	0.729

The best results are highlighted in bold for each metric. AL = Asymmetric loss. BCE = Binary cross entropy loss.

^aThe three best models of each machine learning model.

^bThe overall best model.