RESEARCH

Model-based stratification of progression along the Alzheimer disease continuum highlights the centrality of biomarker synergies

Muhammad Usman Sadiq[1], Kichang Kwak[1], Eran Dayan[1,2*] and for the Alzheimer's Disease Neuroimaging Initiative

Abstract

Background: The progression rates of Alzheimer's disease (AD) are variable and dynamic, yet the mechanisms that contribute to heterogeneity in progression rates remain ill-understood. Particularly, the role of synergies in pathological processes reflected by biomarkers for amyloid-beta ('A'), tau ('T'), and neurodegeneration ('N') in progression along the AD continuum is not fully understood.

Methods: Here, we used a combination of model and data-driven approaches to address this question. Working with a large dataset ($N = 321$ across the training and testing cohorts), we first applied unsupervised clustering on longitudinal cognitive assessments to divide individuals on the AD continuum into those showing fast vs. moderate decline. Next, we developed a deep learning model that differentiated fast vs. moderate decline using baseline AT(N) biomarkers.

Results: Training the model with AT(N) biomarker combination revealed more prognostic utility than any individual biomarkers alone. We additionally found little overlap between the model-driven progression phenotypes and established atrophy-based AD subtypes. Our model showed that the combination of all AT(N) biomarkers had the most prognostic utility in predicting progression along the AD continuum. A comprehensive AT(N) model showed better predictive performance than biomarker pairs (A(N) and T(N)) and individual biomarkers (A, T, or N).

Conclusions: This study combined data and model-driven methods to uncover the role of AT(N) biomarker synergies in the progression of cognitive decline along the AD continuum. The results suggest a synergistic relationship between AT(N) biomarkers in determining this progression, extending previous evidence of A-T synergistic mechanisms.

Keywords: Alzheimer's disease, AT(N) biomarkers, Deep learning, Cognitive phenotypes

*Correspondence: eran_dayan@med.unc.edu
[2] Department of Radiology, UNC-Chapel Hill, Chapel Hill, NC 27599, USA
Full list of author information is available at the end of the article

Background

Alzheimer's disease (AD) is a progressive neurodegenerative disorder which gradually impairs memory, cognition, and other vital functions [1]. Individuals along the AD continuum exhibit markedly heterogeneous progression rates as the disease advances [2, 3]. Both linear and non-linear progression of cognitive decline has been documented in AD [4, 5], with distinct progression profiles found among individuals [2, 3]. Still, the mechanisms
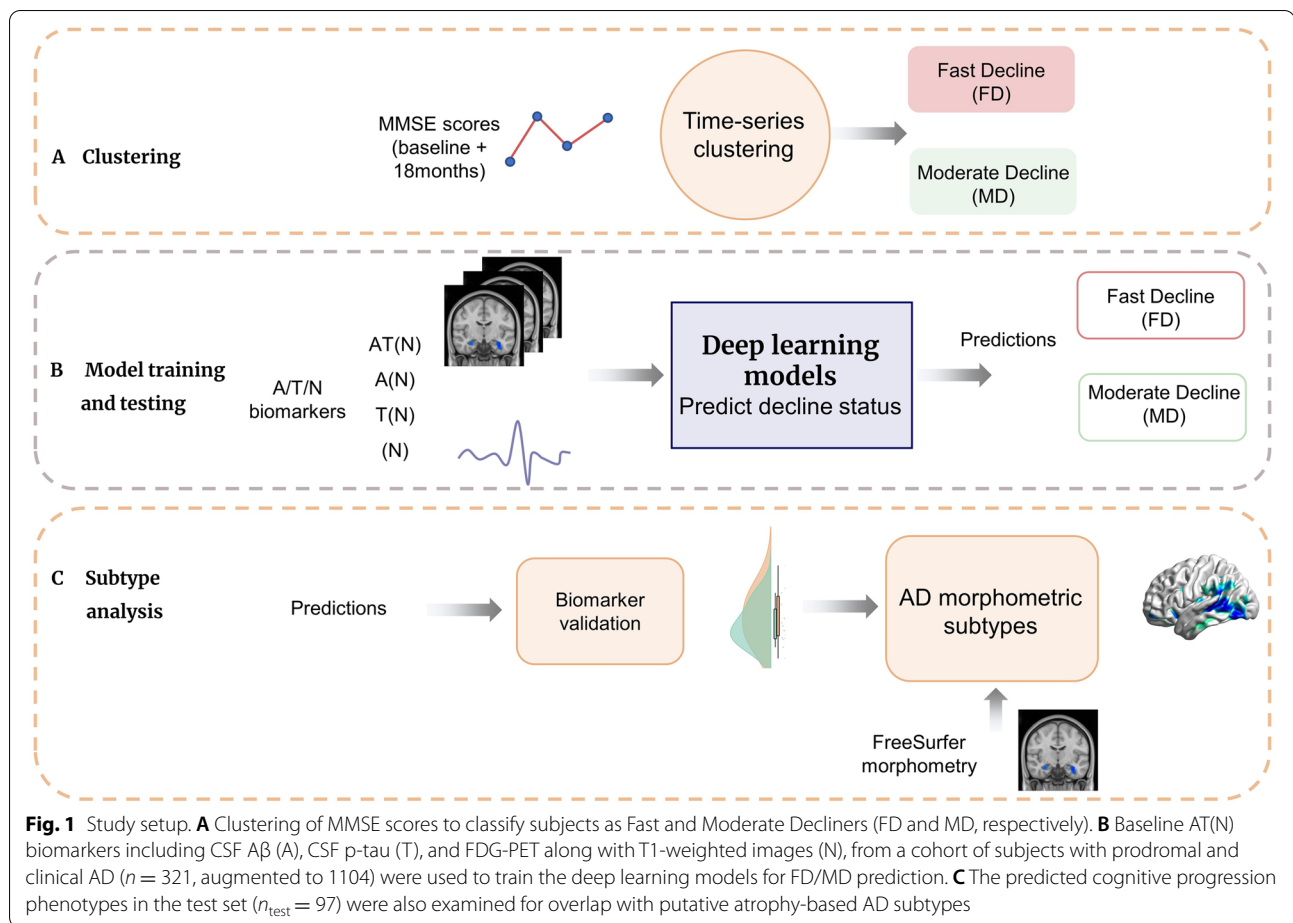
Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 2 of 14

Fig. 1 Study setup. **A** Clustering of MMSE scores to classify subjects as Fast and Moderate Decliners (FD and MD, respectively). **B** Baseline AT(N) biomarkers including CSF Aβ (A), CSF p-tau (T), and FDG-PET along with T1-weighted images (N), from a cohort of subjects with prodromal and clinical AD ($n = 321$, augmented to 1104) were used to train the deep learning models for FD/MD prediction. **C** The predicted cognitive progression phenotypes in the test set ($n_{test} = 97$) were also examined for overlap with putative atrophy-based AD subtypes

that underlie the heterogeneity in AD progression rates remain incompletely understood.

The neuropathological hallmarks of AD are centered around the presence of amyloid-beta (Aβ) plaques and neurofibrillary tangles of hyperphosphorylated tau, which are believed to precede structural neurodegenerative changes in the brain [6, 7]. Links between cognitive decline in AD and biomarker levels for Aβ [8], tau [9], and atrophy/neurodegeneration [10] have been reported in the literature. However, with little exception, studies have focused on individual biomarkers rather than examining their synergies and combined contribution to progressive cognitive decline along the AD continuum. An accurate characterization of the mechanisms leading to heterogeneity in progression rates would nevertheless benefit from considering biomarkers for Aβ ('A'), tau ('T'), and neurodegeneration ('N') together, consistent with the recently proposed AT(N) framework [11, 12]. Yet, combining AT(N) biomarkers in a single model is not trivial, given their complex, non-linear relationships with one another and/or their relationship with cognitive decline [13, 14]. A modeling approach based on deep learning

arises as a natural solution to this problem, given its ability to model complex and non-linear mappings [15, 16]. Deep learning models have emerged as a powerful tool recently in relevant tasks, such as differentiating between individuals with dementia and controls [17, 18], and classifying stable vs. progressive mild cognitive impairment (MCI) [15, 19–21].

In the current study, we propose a model-driven approach, based on AT(N) biomarkers, for stratifying progression rates along the AD continuum and delineating their underlying mechanisms. Notably, this work focuses on heterogeneity of cognitive decline along the AD continuum unlike previous studies where MCI progression was examined [22, 23]. We first employ data-driven clustering of cognitive assessments to define individuals with prodromal or clinical AD as either Fast Decliners (FD) or Moderate Decliners (MD) (Fig. 1A). These progression phenotypes are then used to train, validate, and test a deep learning model using baseline biomarkers for A (CSF Aβ $_{1-42}$), T (CSF p-tau $_{181}$), and N (MRI images and FDG-PET) (Fig. 1B). The model was trained with and without Aβ, tau, and neurodegeneration

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 3 of 14

biomarkers, allowing us to compare the relative contribution of biomarker synergies, particularly amyloid-, and tau-mediated neurodegeneration to progression rates along the AD continuum. We additionally examined the extent to which the cognitive progression phenotypes predicted by our model reflected variation in regional atrophy characteristics (Fig. 1C), commonly used for subtyping AD [24–27]. This allowed us to examine if our model-based framework reflected patterns of neurodegeneration captured by other commonly used approaches.

Methods
time points during 24 months following baseline, (3) had T1-weighted MRI images takenParticipants and data acquisition

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu; clinical trial registration: NCT00106899). The ADNI was launched in 2003 as a public-private partnership, led by the Principal Investigator Michael W. Weiner, M.D. The primary goal of the ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. In the current study, participants from the ADNI-1, ADNI-2/Go, and ADNI-3 cohorts were included if they were (1) diagnosed with AD at baseline or within 1 year of their first diagnosis (i.e., MCI subjects were included if they were later diagnosed as AD), (2) had valid cognitive evaluations (Mini-Mental State Examination (MMSE) scores) performed at a minimum of $t = 4$ time points during 24 months following baseline, (3) had T1-weighted MRI images taken using 3T scanners based on either an inversion recovery-fast spoiled gradient recalled (IR-SPGR) or a magnetization-prepared rapid gradient-echo (MP-RAGE) sequences, and (4) were determined to be amyloid-positive within the study's timeline according to published criteria (CSF Aβ < 976.6 pg/mL or ^{18}F-florbetapir-PET uptake ratio > 1.11). The ^{18}F-florbetapir-PET uptake ratios, provided in ADNI as a derived variable, were calculated by extracting weighted cortical retention means from frontal, cingulate, parietal, and temporal regions, after co-registering the PET and MRI scans. These data were used to calculate standardized uptake value ratios (SUVRs), normalized by a whole cerebellum reference region. SUVRs with a positivity threshold of 1.11 were then identified [28, 29].

In total, 321 unique subjects were identified using these criteria. Out of these, 310 subjects were determined to be amyloid-positive using CSF Aβ cutoff while 11 subjects satisfied the PET uptake ratio criterion. All subjects

provided written informed consent, and the procedures were approved by the Institutional Review Boards of participating centers.

MRI images and their processing

T1-weighted SPGR or MPRAGE images were acquired using 3T scanners (full details of the image acquisition protocols can be found online (http://adni.loni.usc.edu/methods/documents/mri-protocols/). T1-weighted images were used for training of the deep learning models. The cohort with valid MRI and cognitive assessments were split between the training ($n_{train} = 224$) and testing ($n_{test} = 97$) datasets.

CSF Aβ$_{1-42}$ (amyloid-beta) and p-tau$_{181}$ (tau) biomarkers along with MRI + fluorodeoxyglucose (FDG)-PET data (neurodegeneration) were used as the A, T, and N biomarkers respectively in our AT(N)-centered analytical framework. CSF samples used in this study were collected and processed previously (see, [30]; http://adni.loni.usc.edu/methods). CSF Aβ and p-tau were measured with the fully automated Elecsys immunoassay (Roche Diagnostics, Basel, Switzerland) by the ADNI biomarker core (University of Pennsylvania, Philadelphia, PA). Processed (see, http://adni.loni.usc.edu/methods) FDG PET images were averaged, with uptake values from angular, temporal, and posterior cingulate cortices serving as one of our two biomarkers for neurodegeneration (along with MRI) [31]. This average FDG PET was previously obtained using a series of steps to mitigate inter-scanner variability and normalized in spatial resolution and intensity range for further analysis [32]. Each MRI image was standardized to 0 mean and unit standard deviation. Similarly, other AT(N) biomarkers were standardized before being used as input in the deep learning model.

Unsupervised clustering of cognitive measurements

To characterize longitudinal change in cognition, we used 2-year follow-up MMSE scores. In addition, other cognitive assessments over the same duration were used for validation purposes, including the Alzheimer's Disease Assessment Scale, 13-Item Subscale (ADASCog13), Clinical Dementia Rating Sum of Boxes (CDR-SB), and Functional Assessment Questionnaire (FAQ). These tests were administered as described online (http://www.adni-info.org).

We used time-series clustering based on the dynamic time-warping (DTW) method [33] to identify cognitive phenotypes in a data-driven manner. Clustering is typically applied in order to partition a heterogeneous set of samples into more homogeneous clusters based on some similarity measure. When it comes to clustering of time-series data, a DTW-based similarity measure is more widely applicable than the conventional Euclidean distance

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 4 of 14

or spatial distance based measures [34, 35]. DTW is able to find optimal global alignment between sequences of different shapes. The shape-based DTW method is particularly well-suited to dynamic time-series data with potential temporal drift, showing better accuracy than linear models [34, 36]. We used the DTW to cluster the MMSE scores of our cohort using $t = 4$ time points, collected over 2 years from baseline, using Hierarchical Agglomerative Clustering with Ward's linkage [37]. Clustering was repeated with other linkage methods such as Ward1 and the unweighted pair-group method using arithmetic averages (UPGMA) to examine the similarity of cluster labels [38, 39]. Other cognitive assessments such as ADASCog13, CDR-SB, and FAQ were used for validation purposes, testing if the phenotypes based on MMSE scores also differ in other measures of cognition in AD. Further, to determine the optimal number of cognitive decline clusters in our cohort, we used silhouette analysis to compare average silhouette width for $k = 2$, 3, and 4 clusters.

Deep learning model architecture and training

Deep learning models have been extensively used for AD classification [17, 18] and predicting progression of MCI [15, 21, 40]. Deep learning models are typically compared against linear or non-linear Support-Vector Machine (SVM), logistic regression, or random forest classifiers, where SVM has been shown to outperform the latter two [41]. We first calibrated our deep learning model's performance using a similar comparison with SVM. Our deep learning model used a parameter-efficient architecture similar to that previously proposed for classification of MCI [42]. The Parameter-Efficient Network model, designated as PENet, takes a combination of baseline AT(N) biomarkers including MRI images and FDG-PET (*N*), CSF p-tau (*T*), and CSF Aβ (*A*) and learns to predict the subject's cognitive decline status (FD vs MD) using these baseline measurements only. The multi-modal feature extractor implemented in the model uses a series of convolutional blocks, or conv blocks, to process MRI tensors. These conv blocks are composed of a convolutional layer followed by batch normalization and exponential linear unit (ELU) transformation. The model also makes use of separable convolution blocks, or sep-conv blocks, which perform the operation of a convolution block but with far fewer parameters, hence reducing the risk of over-fitting. PENet uses 2 conv blocks followed by 3 sep-conv blocks with increasing number of filters (Fig. 3A). It processes non-imaging biomarkers by dense or FC (fully connected) blocks.

Implementation

Experiments were conducted using python version 3.6. The implementation was developed using the Keras deep learning library with Tensorflow backend. The model was trained on Ubuntu 18.04 on a single Nvidia Tesla V100 GPU with 16G memory, using a batch size of 25 and trained for 50 epochs after which the model showed stable dynamics (Fig. S2). This training was performed using the Stochastic Gradient Descent algorithm with an initial learning rate $= 8 \times 10^{-4}$ and exponential decay with a drop rate $= 0.5$. The FC layers used in the model were regularized using L_2 regularization with penalty coefficient $= 5 \times 10^{-4}$.

Data augmentation and validation framework

The implemented model was trained and validated using 5-fold cross-validation stratified by class phenotypes. All qualifying subjects from ADNI-1, ADNI-2/Go, and ADNI-3 were used in our experiments, yielding a total of $n = 321$ subjects ($n_{train} = 224$, MD = 136, FD = 88; $n_{test} = 97$, MD = 58, FD = 39). To improve model generalizability, we augmented the training dataset through a combination of image rotation (random angle in $[-90°, 90°]$), translation (random shift in $[0, 0.5]$), and flipping operations, resulting in 1104 training images. Special care was taken to use the test dataset only after all steps of augmentation, model selection, and hyperparameter tuning were completed, ensuring no data leakage.

Analysis of atrophy-based AD subtypes

MRI images for the test dataset were processed (http://adni.loni.usc.edu/) using Freesurfer (http://surfer.nmr.mgh.harvard.edu/) to extract region of interest (ROI)-based gray matter (GM) volume. The following processing steps were performed: (1) motion-correction and skull-stripping based on a watershed deformation method [43], (2) image registration to the Talairach brain template, (3) estimation and labeling of gray matter-white matter (GM-WM) boundary using a tessellation step, and (4) registration of volume to an atlas to acquire volume and surface statistics for each ROI. Using these extracted volumes, we investigated the potential association between the model-based cognitive progression phenotypes and atrophy-based AD subtypes, as previously identified [24]. Subtypes were identified using the ratio of hippocampal volume (HV) to cortical total volume (CTV). Following the same procedure as described previously [24], subjects in the test dataset with an HV:CTV ratio above the 75th percentile were identified as belonging to the Hippocampal-Sparing AD (HpSp) subtype, those with HV:CTV ratio below the 25th percentile as belonging to the Limbic-Predominant subtype (LP) and the rest were designated as typical-AD (tAD).

To visualize the spatial extent of atrophy, the subtypes HpSp, LP, and tAD were also contrasted against age-matched controls ($n = 30$) to extract voxel-wise

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 5 of 14

contrast maps using FSL's optimized voxel-based morphometry (VBM) (http://www.fmrib.ox.ac.uk/fsl/) [44]. The FSL VBM processing pipeline involved brain extraction of T1-weighted images followed by segmentation into WM, GM, and CSF volume probability maps. Next, a random subset of each compared cohort was used to create the average study-specific GM template by registration to MNI152 space using the FSL FLIRT tool. This was followed by non-linear registration of all GM images in the native image space to the average GM template. Subsequently, these registered images were smoothed using a full-width half-maximum (FWHM) of 6mm and their voxel-wise GM volumes were contrasted using a general linear model (GLM) formulation. To identify significant differences between the compared groups, non-parametric statistics were performed using the 'randomise' FSL function (5000 permutations) with FWE correction set at $p < 0.05$, based on threshold-free cluster enhancement (TFCE).

Statistical analysis

For comparisons between two groups, unpaired two-sided t tests or Wilcoxon rank-sum test were used. For testing significant differences in MMSE scores of the MD and FD phenotypes, we used the selective inference method [45] implemented in the R 'clusterpval' package, which controls for type I error rate in group comparisons after clustering. To analyze longitudinal changes in ADASCog13, CDR-SB, and FAQ scores, two-way repeated measures ANOVA was used to examine the main effect of time and its interaction with the MD/FD phenotypes.

Results

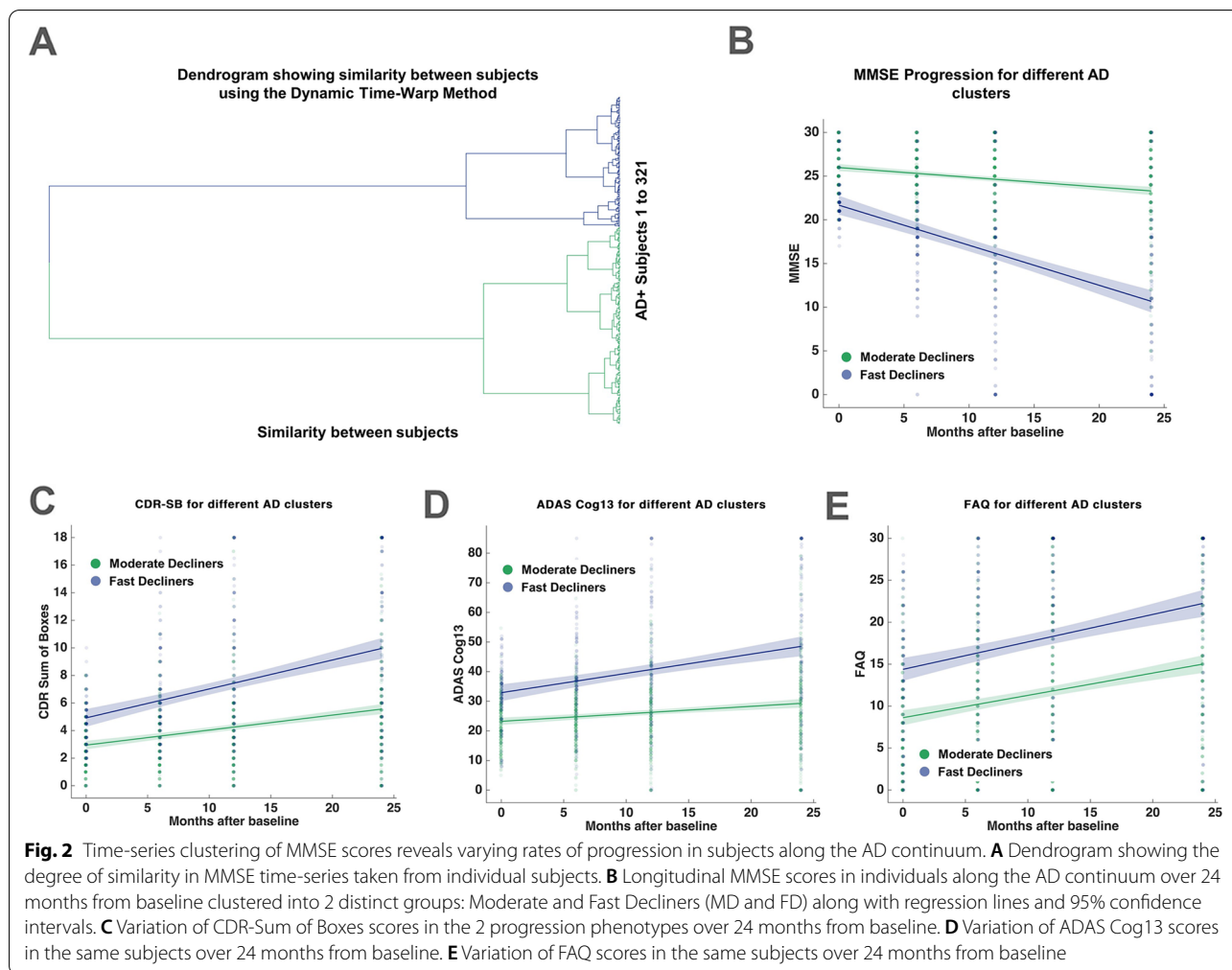### Clustering cognitive assessments according to progression rates

Data used in this study were obtained from the ADNI database [46], combining participants from the ADNI-1, ADNI-2/GO, and ADNI-3 cohorts ($N = 321$; $m = 180$, $f = 141$). We first applied data-driven clustering to derive distinct cognitive progression phenotypes from the sample, considering longitudinal (over 2-year follow-ups) changes in MMSE scores [47]. We used an unsupervised time-series clustering technique based on DTW with Ward's linkage-based agglomerative clustering [33, 37], suitable for shape-based clustering of dynamic time-varying observations [34, 35]. The longitudinal MMSE scores of the entire cohort were clustered to reveal 2 different progression phenotypes identified as moderate (MD: $n = 194$; ages 73.8 ± 7.28; Supplementary Table 1) and fast (FD: $n = 127$; ages 73.2 ± 8.02) decline (Fig. 2A and B). Other linkage methods such as Ward1 [38] and UPGMA resulted in very similar clustering solutions [38, 39]. The 2 clusters did not exhibit any significant differences in age, education, gender, total cortical volume, and

APOE e4 status (all $p > 0.05$; Supplementary Table 1 and Fig. S3). Silhouette analysis, used to determine the optimal number of clusters, resulted in maximal silhouette width for $k = 2$ clusters (Supplementary Table 2). The MD and FD phenotypes showed, as expected, significant differences in MMSE profiles ($p = 6.73 \times 10^{-3}$), revealed using a method developed for post-clustering comparisons [45]. Subjects in the MD and FD phenotypes were also compared for longitudinal changes in ADASCog13, CDR-SB, and FAQ scores. Similar to the MMSE scores, the ADASCog13, CDR-SB, and FAQ scores showed distinct patterns of decline in the different progression phenotypes (Fig. 2C-E). Repeated measures ANOVA for ADASCog13 revealed a significant main effect of time ($F(3, 960) = 48.14$; $p < 2 \times 10^{-10}$) and a significant interaction between time and phenotype ($F(3, 960) = 20.68$; $p < 2 \times 10^{-10}$). Similarly, significant main effects for time as well as significant time by phenotype interactions were observed for CDR-SB (main effect: $F(3, 960) = 85.55$; $p < 2 \times 10^{-12}$; interaction: $F(3, 960) = 24.46$; $p < 2 \times 10^{-12}$), and FAQ scores (main effect: $F(3, 960) = 75.49$; $p < 2 \times 10^{-13}$; interaction: $F(3, 960) = 3.13$; $p = 0.024$).

Deep learning model predictions and role of AT(N) biomarkers in progression along AD continuum

The FD and MD phenotypes identified by clustering were next used for model training, where a combination of baseline AT(N) biomarkers including MRI images and FDG-PET (*N*), CSF p-tau $_{181}$ (*T*), and CSF Aβ $_{1-42}$ (*A*) were used as input to the model. We constructed a parameter-efficient deep learning architecture similar to that previously proposed [42]. This architecture uses fewer parameters than most other models used in similar tasks with state-of-the-art performance and is well-suited for data-limited applications. The overall architecture makes use of 2 convolutional blocks followed by 3 separable convolution blocks with progressively increasing number of filters (Fig. 3A; See the "Methods" section for more details). In the model, non-imaging biomarkers are processed by dense or FC (fully connected) blocks. Finally, extracted features from MRI and biomarkers are combined through FC layers to generate predictions. The implemented model was trained and validated using 5-fold cross-validation stratified by cognitive phenotypes. All subjects meeting our criteria for analysis from ADNI-1, ADNI-2/Go, and ADNI-3 were used in the experiments, yielding a total of $n = 321$ subjects ($n_{train} = 224$, MD = 136, FD = 88; $n_{test} = 97$, MD = 58, FD = 39).
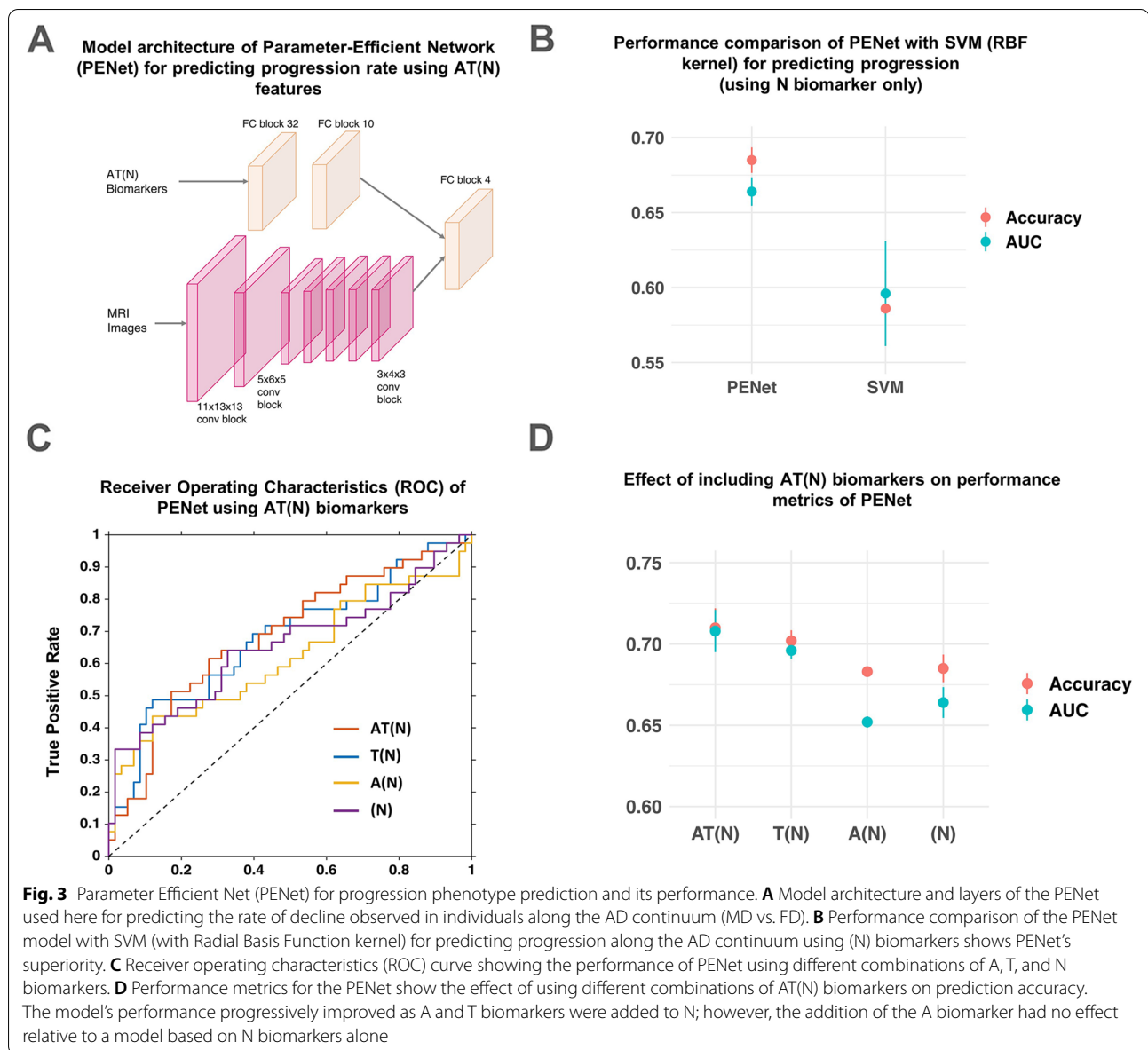
We first evaluated the performance of a basic PENet model trained using N features alone. The model showed significantly better performance than a random classifier when compared using DeLong's test ($p = 0.014$; Fig. S1)

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 6 of 14

Fig. 2 Time-series clustering of MMSE scores reveals varying rates of progression in subjects along the AD continuum. **A** Dendrogram showing the degree of similarity in MMSE time-series taken from individual subjects. **B** Longitudinal MMSE scores in individuals along the AD continuum over 24 months from baseline clustered into 2 distinct groups: Moderate and Fast Decliners (MD and FD) along with regression lines and 95% confidence intervals. **C** Variation of CDR-Sum of Boxes scores in the 2 progression phenotypes over 24 months from baseline. **D** Variation of ADAS Cog13 scores in the same subjects over 24 months from baseline. **E** Variation of FAQ scores in the same subjects over 24 months from baseline

[48]. The PENet model was then also compared with an SVM classifier, which used a radial-basis function kernel and N features as input. The PENet outperformed the SVM model (Fig. 3B), achieving an accuracy of 0.685 (area under the curve, AUC = 0.66).

Next, different combinations of AT(N) biomarkers were used in the PENet model to test their prognostic value and evaluate the role of biomarker synergies in progression along the AD continuum. This method of assessing the predictive successes of different biomarker combinations was preferred over linear or non-linear regression-based methods because of the possibility of complex, higher order synergies between AT(N) biomarkers, especially when 3D voxel-level data is included [14, 49]. Including all AT(N) biomarkers in the model yielded the best classification accuracy (0.710 ± 0.024), followed by T(N) (0.702 ± 0.013), N (0.685 ± 0.017), and A(N) (0.683 ± 0.006) (Fig. 3C, D). Individual A and T biomarkers, when used to predict progression phenotype using a logistic regression model, yielded only 59.7% and
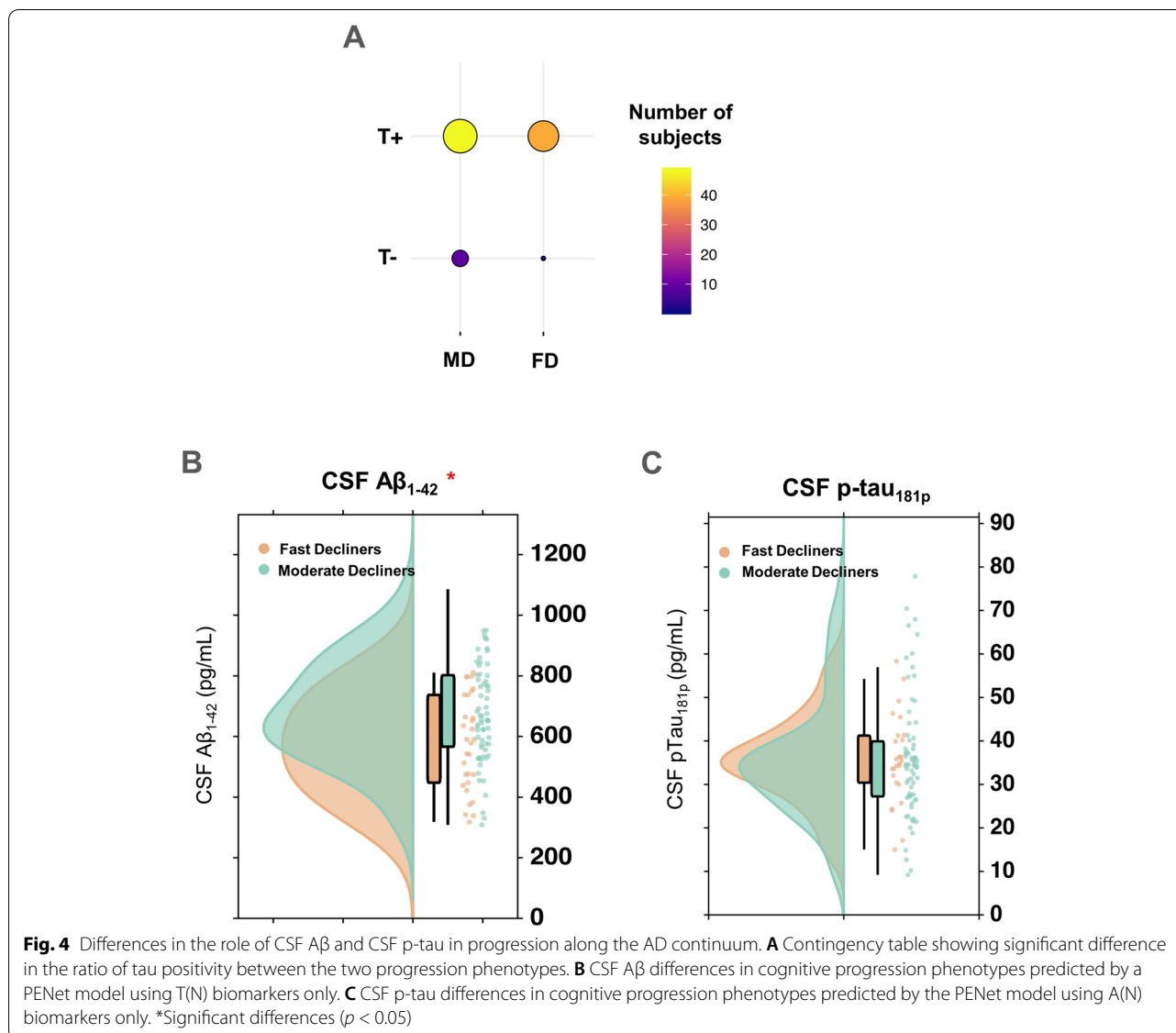
56.7% classification accuracies (Supplementary Table 4). Thus, a comprehensive model based on A, T, and N was more accurate at predicting AD progression phenotypes than biomarker pairs (A(N) and T(N)) or individual biomarkers (A, T, or N). Moreover, relative to a model based on N features alone, the addition of A features resulted in slightly worse accuracy, while the addition of T features resulted in 1.7% improvement. In a secondary analysis, we also investigated the effect of including a proxy for brain reserve and excluding FDG-PET from the model's inputs. In the first experiment, intracranial volume (ICV), a common proxy for brain reserve [50], was used as input to the model together with N biomarkers (MRI and FDG-PET). In the second experiment, FDG-PET was removed from the N biomarkers so that only MRI images were used as N inputs. The two experiments resulted in classification accuracies of 0.686 ($\pm$ 0.021) and 0.677 ($\pm$ 0.010) respectively, compared to that observed when using N biomarkers without these changes (0.685 $\pm$ 0.017) (Supplementary Table 5).

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 7 of 14

Fig. 3 Parameter Efficient Net (PENet) for progression phenotype prediction and its performance. **A** Model architecture and layers of the PENet used here for predicting the rate of decline observed in individuals along the AD continuum (MD vs. FD). **B** Performance comparison of the PENet model with SVM (with Radial Basis Function kernel) for predicting progression along the AD continuum using (N) biomarkers shows PENet's superiority. **C** Receiver operating characteristics (ROC) curve showing the performance of PENet using different combinations of A, T, and N biomarkers. **D** Performance metrics for the PENet show the effect of using different combinations of AT(N) biomarkers on prediction accuracy. The model's performance progressively improved as A and T biomarkers were added to N; however, the addition of the A biomarker had no effect relative to a model based on N biomarkers alone

We next tested whether the predicted cognitive progression phenotypes differed in tau positivity (where T+ was defined as CSF p-tau > 21.8 pg/ml, [12]). We found significant differences in the proportion of tau positivity between the MD and FD groups (χ^2= 4.48, p = 0.034; Fig. 4A). The differences between the predictive value of CSF p-tau and CSF Aβ were then further validated, by testing the degree of redundancy in their predictive value. To do this, the MD and FD phenotypes as predicted by A(N) biomarkers were tested for significant differences in CSF p-tau. Similarly, the two phenotypes as predicted by T(N) biomarkers were tested for significant differences in CSF Aβ. Significant differences were only found between CSF Aβ levels in the

model where the MD/FD phenotypes were predicted using T(N) biomarkers (p = 0.02). No significant differences were observed between CSF p-tau levels in the model where the predicted MD/FD labels were based on A(N) biomarkers (Fig. 4B, C). Thus, the combination of T(N) features could account for the variance in CSF Aβ between the two phenotypes whereas the combination of A(N) features could not account similarly for the variation in CSF p-tau.

The moderate and fast decliners were distributed according to clinical diagnosis as follows: MD (MCI = 123, AD = 71) and FD (MCI = 22, AD = 105). Thus, there were more MCI subjects in the MD group, and more AD subjects in the FD group, respectively.

Fig. 4 Differences in the role of CSF Aβ and CSF p-tau in progression along the AD continuum. **A** Contingency table showing significant difference in the ratio of tau positivity between the two progression phenotypes. **B** CSF Aβ differences in cognitive progression phenotypes predicted by a PENet model using T(N) biomarkers only. **C** CSF p-tau differences in cognitive progression phenotypes predicted by the PENet model using A(N) biomarkers only. *Significant differences ($p < 0.05$)

Therefore, to examine if the model was only learning MCI/AD characteristics instead of reflecting the heterogeneity of cognitive decline along the AD continuum, we tested our model trained on MD/FD classification in the task of detecting the MCI/AD clinical diagnosis. Using the 5-fold cross-validation setup used in previous experiments, the model predicted the MCI/AD diagnosis with 0.584 ± 0.03 accuracy, showing substantially worse performance than that observed in the task of MD/FD classification. This demonstrates the distinct nature of the MD/FD and MCI/AD classification problems.

Association of cognitive progression phenotypes with AD subtypes

Using extracted GM volumes for subjects in the test set (see the "Methods" section), we next investigated the potential association between the progression phenotypes identified here and atrophy-based AD subtypes as previously identified [24]. The objective of this analysis was to examine if the cognitive progression phenotypes predicted by our model reflected AD subtypes captured solely by patterns of neurodegeneration. The atrophy-based AD subtypes were defined based on the HV:CTV ratio. This resulted in a total of 26, 24, and 47 subjects assigned to the HpSp, LP, and tAD subtypes, respectively. First, to validate the presence of previously identified atrophy patterns in these subtypes, the spatial extent of their regional atrophy was visualized by contrasting them against data from age-matched controls ($n = 30$; Fig. 5A). These maps were obtained using the FSL VBM approach [44]. Consistent with previously observed atrophy patterns [51], the HpSp subtype was manifested in
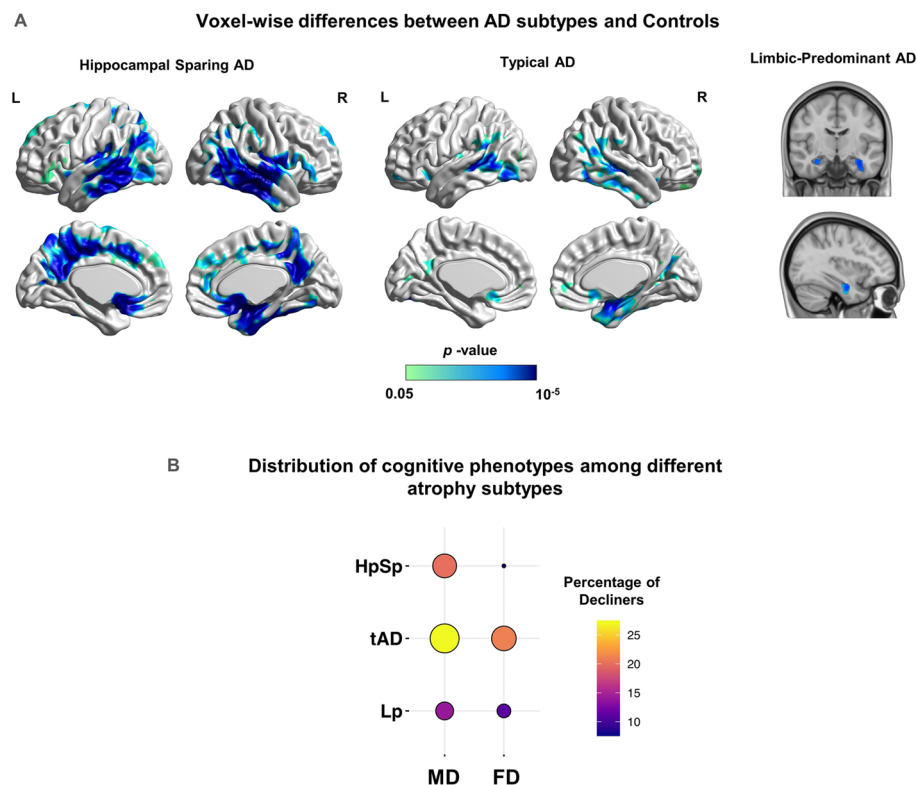
Fig. 5 Association between progression phenotypes predicted by the deep learning model and AD subtypes identified by morphometric analysis. **A** Voxel-wise differences between the different AD subtypes and a group of age-matched healthy controls (*n*=30) reveal the regional atrophy patterns in these subtypes. The HpSp subtype (associated mostly with the FD phenotype) was manifested in widespread atrophy spread to the bilateral temporoparietal cortex, precuneus, and posterior cingulate regions while the LP subtype (corresponding mostly with the MD phenotype) displayed voxel-wise differences only in limbic regions. Voxel-wise differences were calculated using FSL-VBM with FWE correction set at *p* < 0.05, based on the threshold-free cluster enhancement (TFCE) statistic. **B** Contingency table showing the association between AD subtypes and the progression phenotypes predicted by the deep learning model. Overall, the different subtypes did not exhibit a significantly different distribution between the two progression phenotypes (χ^2 test, *p* = 0.497)

atrophy spread to bilateral temporoparietal cortex, precuneus, and posterior cingulate regions. On the other hand, the LP subtype displayed voxel-wise differences only in limbic regions. Next, we determined if the atrophy-based AD subtypes overlapped with the predicted cognitive progression phenotypes (MD/FD) identified via our modeling approach. We found no significant differences in the distribution of the different atrophy subtypes among the two cognitive progression phenotypes (χ^2= 1.39, *p* = 0.497; Fig. 5B). Altogether, the HpSp subtype was more prevalent in the MD compared to the FD phenotype (19/7), while the tAD (26/21) and LP (13/11) subtypes were more equally distributed among the MD and FD phenotypes.

Discussion
The mechanisms that underlie the marked heterogeneity in AD progression are to date incompletely understood. In the present work, we used a combination of data and model-driven methods to investigate heterogeneity along the AD continuum and tease apart the contribution of AT(N) biomarkers and their synergies to this progression. We first demonstrated that progression along the AD continuum varied considerably between individuals by applying unsupervised clustering on longitudinal cognitive assessments. Next, we used a parameter-efficient deep neural network (PENet) to predict the different cognitive progression phenotypes using baseline AT(N) biomarkers. Our model showed that the combination of all AT(N) biomarkers had the most prognostic utility in predicting progression along the AD continuum. A comprehensive AT(N) model showed better predictive performance than biomarker pairs (A(N) and T(N)) and individual biomarkers (A, T, or N). Finally, we report that the cognitive progression phenotypes did not overlap with previously established AD subtypes, defined solely based on patterns of neurodegeneration (atrophy). Altogether, our findings highlight a central role for

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 10 of 14

concurrent AT(N) mechanisms which determine the rate of cognitive decline along the AD continuum.

Comparison of age, gender, education, total cortical volume, and *APOE e4* status at baseline revealed no significant differences between the various cognitive phenotypes identified using DTW clustering. This is noteworthy because of the concern that unsupervised clustering may capture variance not necessarily relevant to disease expression patterns, but rather may reflect confounds such as age, gender, or education. While baseline differences in MMSE scores between the phenotypes may have arisen due to differences in disease duration which can only be estimated indirectly in neurodegenerative disease [52], the accentuation of these differences with time indicates that the phenotypes declined at a significantly different rate. Our findings of discrete cognitive progression phenotypes are consistent with previous work on identifying distinct cognitive subtypes in AD [3, 53]. However, our clustering approach is specifically suitable for finding optimal global alignment between time-series data. Since cognitive phenotypes have been suggested to progress at different speeds in various stages of the disease (linearly or non-linearly) [4, 5], more complex modeling methods such as DTW are needed to capture the shape of the progression curve, taking into account potential temporal shifts.

The cognitive phenotypes identified by our analysis also differed significantly in ADASCog13, CDR-SB, and FAQ scores at the 2 years mark. Consistent with our results, differences in ADASCog13, CDR-SB, and FAQ scores were found among some or all of the cognitive subtypes identified in previous studies [24, 54, 55]. In other studies where a more detailed profiling of memory and cognitive domains was included, atypical subtypes were also identified based on visuospatial functioning and language impairment [56, 57]. Moreover, our finding of two distinct cognitive phenotypes is supported by previous studies [3, 58], which have robustly identified two cognitive clusters across multiple AD dementia cohorts, using a data-driven approach.

The MD and FD phenotypes identified via clustering were treated as labels in a model-driven classification and prediction of subjects' cognitive progression phenotype using baseline N biomarkers. Comparison of the deep learning model with an SVM classifier showed better performance for the former model, similar to results published in similar tasks such as MCI classification [41, 59]. Using the PENet deep learning model and neurodegeneration biomarkers only (T1-weighted images and FDG-PET) resulted in baseline accuracy of 0.685, which progressively increased to 0.710 when adding additional T/N biomarkers. Critically, a comprehensive/complete AT(N) model showed more predictive power

than biomarker pairs (A(N) and T(N)) and individual biomarkers (A, T, or N), suggesting that complex synergies between all three biomarker types underlie the progressions of cognitive decline along the AD continuum. This finding extends the emerging evidence from animal model and human studies on synergies between Aβ and tau [60, 61]. For example, injection of $Aβ_{42}$ fibrils into the brains of P301L mutant tau transgenic mice caused a substantial increase in the numbers of neurofibrillary tangles both near the injection site and in regions projecting to it [62]. Evidence pointing to synergies between Aβ and tau also comes from studies in humans [63, 64]. Tau spread outside entorhinal cortex is enhanced by Aβ deposition in cognitively normal older adults [65]. Moreover, significant interactions between CSF Aβ and CSF p-tau affecting brain structure were reported in preclinical AD [66]. Similarly, findings based on PET imaging demonstrate that both Aβ and tau underlie memory decline in preclinical AD [64]. Altogether, previous evidence on synergies between Aβ and tau have mostly originated from studies in cognitively normal or preclinical AD populations [64, 67]. Our study extends this work, examining progression along the AD continuum. Moreover, our results suggest that synergies between all three biomarker types (Aβ, tau, and neurodegeneration) underlie the progression of cognitive decline along the AD continuum. Our deep learning model and its reliance on AT(N) features is also distinct from previous machine learning models where multi-modal AD classification was based on N features (e.g., MRI and FDG) [40, 68] or combined cognitive or demographic variables [42, 69]. Further, previous machine learning based models have been mostly deployed to distinguish AD/CN or predict progression of MCI (see [15, 41] for a review), whereas our investigation focused on cognitive progression phenotypes in the AD continuum.

Our results reveal that the addition of A (Aβ) biomarkers to the model resulted in effectively no increase in accuracy, while the inclusion of T (tau) biomarkers resulted in an improvement of 1.7% in accuracy. This role of T biomarkers in progression along the AD continuum was further highlighted by the significant differences in tau positivity found between the MD and FD phenotypes. Thus, evidence suggests that CSF p-tau is a stronger determinant of progression rate along AD continuum than CSF Aβ, consistent with previous reports showing that tau-mediated neurodegeneration mechanisms result in heterogeneous AD progression [9, 70]. This observation of tau-associated progression along the AD continuum is also in agreement with previous studies where MCI progression was linked to elevated CSF p-tau [71, 72]. Further, the addition of ICV as an N biomarker in the model did not result in a meaningful change in

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 11 of 14

classification accuracy (+0.1%), likely due to the covariance between ICV and voxel-level N biomarkers. Similarly, the exclusion of FDG-PET from the model's N biomarkers had a marginal effect on accuracy (−0.8%).

Since the inclusion of more variables did not always improve the prediction accuracy of the model, the performance differences between the model based on AT(N) inputs and those obtained when using individual inputs cannot be explained solely by the number of input variables in the model. Additionally, the question of whether Aβ, tau, and neurodegeneration biomarkers have an additive or interactive relationship can improve our understanding of the disease. Our results indicate that performance gains associated with the addition of the A biomarker to a model trained on T(N) inputs is different from the improvement gained by adding A to (N) inputs only. In other words, performance improvement due to A is a function of whether T is included in the model's inputs, which suggests that the A and T synergies found in our model are non-additive in nature. We, however, acknowledge that more formal ways of testing interaction vs. additive effects such as partial dependency plots should be employed in future studies to investigate the nature of biomarker synergies in progression along the AD continuum.

The deep learning model deployed here demonstrated good discriminative performance, where special care was taken to avoid different sources of data leakage previously identified [41]. Deep learning models have been extensively used for predicting progression of MCI to AD [15, 19, 21], along with other related tasks such as AD classification with missing data [73] and early detection of AD [74]. However, little is known about their utility in identifying cognitive progression phenotypes using the AT(N) framework. Employing this approach allowed us to investigate the complex and potentially non-linear relationships between these biomarkers. Furthermore, the shape-based DTW clustering of cognitive assessments, as applied here, allowed us to more optimally capture progression rates in AD, as it is specifically designed for modeling time-varying data [36].

We tested whether the cognitive progression phenotypes were captured by an established subtyping approach [24] based solely of patterns of neurodegeneration. We found no significant differences in the distribution of AD subtypes among the two cognitive progression phenotypes. Thus, the model-based phenotypes identified here, may not be readily detectable using atrophy-based methods. Data-driven studies on AD subtyping revealed neurodegeneration patterns similar to those found here [24–26], but atrophy-based methods do not always result in distinct cognitive progression phenotypes [27]. Additional work is needed to better reconcile atrophy and cognitive-based subtyping of AD and its progression.

Limitations
Several limitations should be noted. First, we acknowledge that any study investigating the effects of AT(N) biomarkers in AD should ideally test longitudinal changes in these biomarkers in the same cohort. However, this was not applicable in the current study due to missing data in several of the biomarkers. Second, while being beyond the scope of the current study, an examination of the spatiotemporal characteristics of amyloid and tau deposition using PET-based markers can provide useful information about the progression of cognitive decline along the AD continuum. Future work focusing on spatiotemporal changes in the synergy between biomarkers for Aβ, tau, and neurodegeneration as it relates to progression rates is thus warranted.

Conclusions
To conclude, our study combined data and model-driven methods to uncover the role of AT(N) biomarkers in the progression of cognitive decline along the AD continuum. The results converge to support a more complex, synergistic relationship between AT(N) biomarkers in determining this progression. Our findings further demonstrate the utility of using modeling approaches to study the complex multifaceted mechanisms that underlie disease progression in AD.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13195-021-00941-1.

Additional file 1.

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 12 of 14

Availability of data and materials
Data used in this study were obtained from the ADNI database (adni.loni.usc. edu), excluded in specific cases according to pre-established exclusion criterion determined by study design constraints such as MRI image resolution, availability of longitudinal data, subject's amyloid-beta levels, etc.

Declarations

Ethics approval and consent to participate
ADNI investigators obtained ethics approval from the local ethical committees of all involved sites. Access to all ADNI data was granted to us after registration to ADNI (https://adni.loni.usc.edu) and compliance with the data usage agreement. All work complied with ethical regulations for work with human participants.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Biomedical Research Imaging Center (BRIC), UNC-Chapel Hill, Chapel Hill, NC 27599, USA. [2]Department of Radiology, UNC-Chapel Hill, Chapel Hill, NC 27599, USA.

References

1. Sloane PD, Zimmerman S, Suchindran C, Reed P, Wang L, Boustani M, et al. The public health impact of Alzheimer's disease, 2000–2050: potential implication of treatment advances. Annu. Rev. Public Health. 2002:213–31 [cited 2021 Feb 10] Available from: https://pubmed.ncbi. nlm.nih.gov/11910061/.
2. Eldholm RS, Barca ML, Persson K, Knapskog AB, Kersten H, Engedal K, et al. Progression of Alzheimer's disease: a longitudinal study in Norwegian memory clinics. J Alzheimer's Dis. 2018;61:1221–32 [cited 2021 Feb 9] Available from: https://pubmed.ncbi.nlm.nih.gov/29254085/.
3. Scheltens NME, Tijms BM, Koene T, Barkhof F, Teunissen CE, Wolfsgruber S, et al. Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts. Alzheimer's Dement. 2017;13:1226–36 [cited 2021 Feb 9] Available from: /pmc/articles/PMC5857387/?report=abstract.
4. Stern Y, Liu X, Albert M, Brandt J, Jacobs DM, Castillo-Castaneda CD, et al. Application of a growth curve approach to modeling the progression of Alzheimer's disease. Journals Gerontol Ser A Biol Sci Med Sci. 1996;51A:M179–84. [cited 2021 Feb 9] Available from: https://academic.oup.com/biome dgerontology/article-lookup/doi/10.1093/gerona/51A.4.M179
5. Doody RS, Pavlik V, Massman P, Rountree S, Darby E, Chan W. Predicting progression of Alzheimer's disease. Alzheimer's Res Ther. 2010;2:2. [cited 2021 Mar 31] Available from: http://alzres.biomedcentral.com/articles/10. 1186/alzrt25
6. Meyer PF, Pichet Binette A, Gonneaud J, Breitner JCS, Villeneuve S. Characterization of Alzheimer disease biomarker discrepancies using cerebrospinal fluid phosphorylated tau and AV1451 positron emission tomography. JAMA Neurol. 2020;77:508–16 [cited 2021 Apr 26] Available from: http:// adni.loni.usc.edu.
7. Deture MA, Dickson DW. The neuropathological diagnosis of Alzheimer's disease. Mol. Neurodegener. 2019:1–18 [cited 2021 Mar 31] Available from: https://doi.org/10.1186/s13024-019-0333-5.
8. Haass C, Selkoe DJ. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. Nat. Rev. Mol. Cell Biol. 2007:101–12 [cited 2021 Feb 11] Available from: www.nature.com/revie ws/molcellbio.
9. Ballatore C, Lee VMY, Trojanowski JQ. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. Nat. Rev. Neurosci. 2007:663–72 [cited 2021 Feb 11] Available from: www.nature.com/reviews/neuro.
10. Fox NC, Warrington EK, Freeborough PA, Hartikainen P, Kennedy AM, Stevens JM, et al. Presymptomatic hippocampal atrophy in Alzheimer's disease. Brain. 1996;119:2001–7. [cited 2021 Feb 11] Available from: https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/119.6. 2001
11. Jack CR, Bennett DA, Blennow K, Carrillo MC, Feldman HH, Frisoni GB, et al. A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology. 2016:539–47 [cited 2021 Mar 2] Available from: /pmc/articles/PMC4970664/.
12. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimer's Dement. 2018:535–62 [cited 2021 Apr 28] Available from: https://pubmed.ncbi.nlm.nih.gov/29653606/.
13. Williams JH, Wilcock GK, Seeburger J, Dallob A, Laterza O, Potter W, et al. Non-linear relationships of cerebrospinal fluid biomarker levels with cognitive function: an observational study. Alzheimer's Res Ther. 2011;3:5. [cited 2021 Mar 31] Available from: http://alzres.biomedcentral.com/artic les/10.1186/alzrt64
14. Popescu SG, Whittington A, Gunn RN, Matthews PM, Glocker B, Sharp DJ, et al. Nonlinear biomarker interactions in conversion from mild cognitive impairment to Alzheimer's disease. Hum Brain Mapp. 2020;41:4406–18 [cited 2021 Mar 31] Available from: https://onlinelibrary.wiley.com/doi/ 10.1002/hbm.25133.
15. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–48 [cited 2021 Mar 31] Available from: http:// www.annualreviews.org/doi/10.1146/annurev-bioeng-071516-044442.
16. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med. Image Anal. 2017:60–88.
17. Islam J, Zhang Y. Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Informatics. 2018;5 [cited 2021 Feb 9] Available from: /pmc/articles/PMC6170939/.
18. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF, Weiner M, et al. Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. Sci Rep. 2018;8 [cited 2021 Feb 9] Available from: https://pubmed.ncbi.nlm.nih. gov/29632364/.
19. Zhang D, Wang Y, Zhou L, Yuan H, Shen D. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. 2011;55:856–67.
20. Kwak K, Niethammer M, Giovanello KS, Styner M, Dayan E. Differential Role for hippocampal subfields in Alzheimer's disease progression revealed with deep learning. Cereb Cortex. 2021; [cited 2021 Aug 15] Available from: https://pubmed.ncbi.nlm.nih.gov/34322704/.
21. Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front Aging Neurosci. 2019; 11:220.[cited 2021 Feb 9] Available from: https:// www.frontiersin.org/article/10.3389/fnagi.2019.00220/full.
22. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia - meta-analysis of 41 robust inception cohort studies. Acta Psychiatr Scand. 2009;119:252–65.[cited 2021 Feb 10] Available from: http://doi.wiley.com/10.1111/j.1600-0447.2008.01326.x
23. Duara R, Loewenstein DA, Greig MT, Potter E, Barker W, Raj A, et al. Pre-MCI and MCI: Neuropsychological, clinical, and imaging features and progression rates. Am J Geriatr Psychiatry. 2011;19:951–60 [cited 2021 Feb 10] Available from: /pmc/articles/PMC3175279/.

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 13 of 14

24. Risacher SL, Anderson WH, Charil A, Castelluccio PF, Shcherbinin S, Saykin AJ, et al. Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline. Neurology. 2017;89:2176–86 [cited 2021 Feb 6] Available from: /pmc/articles/PMC5696639/?report=abstract.

25. Ferreira D, Shams S, Cavallin L, Viitanen M, Martola J, Granberg T, et al. The contribution of small vessel disease to subtypes of Alzheimer's disease: a study on cerebrospinal fluid and imaging biomarkers. Neurobiol Aging. 2018;70:18–29 [cited 2021 Feb 8] Available from: https://pubmed.ncbi.nlm.nih.gov/29935417/.

26. Park JY, Na HK, Kim S, Kim H, Kim HJ, Seo SW, et al. Robust identification of Alzheimer's disease subtypes based on cortical atrophy patterns. Sci Rep. 2017;7 [cited 2021 Feb 9] Available from: /pmc/articles/PMC5343676/.

27. Poulakis K, Pereira JB, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, et al. Heterogeneous patterns of brain atrophy in Alzheimer's disease. Neurobiol Aging. 2018;65:98–108 [cited 2021 Feb 9] Available from: https://pubmed.ncbi.nlm.nih.gov/29455029/.

28. Landau SM, Marks SM, Mormino EC, Rabinovici GD, Oh H, O'Neil JP, et al. Association of lifetime cognitive engagement and low β-amyloid deposition. Arch Neurol. 2012 ;69:623–9. [cited 2021 Oct 23] Available from: https://pubmed.ncbi.nlm.nih.gov/22271235/

29. Landau SM, Breault C, Joshi AD, Pontecorvo M, Mathis CA, Jagust WJ, et al. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. J Nucl Med. 2013;54:70–7 [cited 2021 Feb 2] Available from: https://pubmed.ncbi.nlm.nih.gov/23166389/.

30. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009;65:403–13 [cited 2021 Feb 2] Available from: /pmc/articles/PMC2696350/?report=abstract.

31. Toussaint PJ, Perlbarg V, Bellec P, Desarnaud S, Lacomblez L, Doyon J, et al. Resting state FDG-PET functional connectivity as an early biomarker of Alzheimer's disease using conjoint univariate and independent component analyses. Neuroimage. 2012;63:936–46.

32. WJ J, SM L, LM S, JQ T, RA K, EM R, et al. Relationships between biomarkers in aging and dementia. Neurology. 2009;73:1193–1199. [cited 2021 Oct 22] Available from: https://pubmed.ncbi.nlm.nih.gov/19822868/

33. Petitjean F, Ketterlin A, Gançarski P. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognit. 2011;44:678–93.

34. Aach J, Church GM. Aligning gene expression time series with time warping algorithms. Bioinformatics. 2001;17:495–508. [cited 2021 Feb 3] Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/17.6.495

35. Kruskal JB. An overview of sequence comparison: time warps, string edits, and macromolecules*. 1983. Available from: https://epubs.siam.org/page/terms

36. Aghabozorgi S, Shirkhorshidi AS, Wah Y. Time-series clustering – a decade review. 2015 [cited 2021 Feb 9]; Available from: http://dx.doi.org/https://doi.org/10.1016/j.is.2015.04.007

37. Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc. 1963;58:236–44.

38. Murtagh F, Legendre P. Ward's Hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J Classif. 2014;31:274–95 [cited 2021 Jul 14] Available from: www.clustan.com.

39. Montero P, Vilar JA. TSclust: an R package for time series clustering. J Stat Softw. 2014;62:1–43.

40. Il SH, Lee SW, Shen D. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage. 2014;101:569–82.

41. Wen J, Thibeau-Sutre E, Diaz-Melo M, Samper-González J, Routier A, Bottani S, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. Med Image Anal2020;63:101694.

42. Spasov S, Passamonti L, Duggento A, Liò P, Toschi N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. Neuroimage. 2019;189:276–87.

43. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002;33:341–55.

44. Good CD, Johnsrude IS, Ashburner J, Henson RNA, Friston KJ, Frackowiak RSJ. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage. 2001;14:21–36 [cited 2021 Feb 3] Available from: https://pubmed.ncbi.nlm.nih.gov/11525331/.

45. Gao LL, Bien J, Witten D. Selective inference for hierarchical clustering. arXiv. 2020 [cited 2021 Apr 1]; Available from: http://arxiv.org/abs/2012.02936

46. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging. 2008:685–91.

47. Folstein MF, Robins LN, Helzer JE. The Mini-Mental State Examination. Arch. Gen. Psychiatry. 1983:812 [cited 2021 Mar 31] Available from: https://jamanetwork.com/journals/jamapsychiatry/fullarticle/493108.

48. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a non-parametric approach. Biometrics. 1988;44:837.

49. Gispert JD, Rami L, Sánchez-Benavides G, Falcon C, Tucholka A, Rojas S, et al. Nonlinear cerebral atrophy patterns across the Alzheimer's disease continuum: impact of APOE4 genotype. Neurobiol Aging. 2015;36:2687–701.

50. Groot C, van Loenhoud AC, Barkhof F, van Berckel BNM, Koene T, Teunissen CC, et al. Differential effects of cognitive reserve and brain reserve on cognition in Alzheimer disease. Neurology. 2008;90(2):e149-56. https://doi.org/10.1212/WNL.0000000000004802.

51. Ossenkoppele R, Lyoo CH, Sudre CH, Westen D, Cho H, Ryu YH, et al. Distinct tau PET patterns in atrophy-defined subtypes of Alzheimer's disease. Alzheimer's dement. 2020;16:335–44 [cited 2021 Apr 5] Available from: https://onlinelibrary.wiley.com/doi/abs/10.1016/j.jalz.2019.08.201.

52. Hall CB, Lipton RB, Sliwinski M, Stewart WF. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. Stat Med. 2000;19:1555–1566. [cited 2021 Apr 28] Available from: https://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-0258(20000615/30)19:11/12%3C1555::AID-SIM445%3E3.0.CO;2-3

53. Schmidt C, Wolff M, Weitz M, Bartlau T, Korth C, Zerr I. Rapidly progressive alzheimer disease. Arch. Neurol. 2011:1124–30 [cited 2021 Feb 9] Available from: https://jamanetwork.com/.

54. Haaksma ML, Calderón-Larrañaga A, Olde Rikkert MGM, Melis RJF, Leoutsakos J-MS. Cognitive and functional progression in Alzheimer disease: a prediction model of latent classes. Int J Geriatr Psychiatry. 2018;33:1057–64 [cited 2021 Feb 9] Available from: http://doi.wiley.com/10.1002/gps.4893.

55. Bernick C, Cummings J, Raman R, Sun X, Aisen P. Age and rate of cognitive decline in Alzheimer disease: implications for clinical trials. Arch Neurol. 2012;69:901–5 [cited 2021 Feb 9] Available from: http://www.r-project.

56. Crane PK, Trittschuh E, Mukherjee S, Saykin AJ, Sanders RE, Larson EB, et al. Incidence of cognitively defined late-onset Alzheimer's dementia subgroups from a prospective cohort study. Alzheimer's Dement. 2017;13:1307–16 [cited 2021 Feb 9] Available from: https://pubmed.ncbi.nlm.nih.gov/28623677/.

57. Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C. Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods. Biol. Psychiatry. 2020. 70–82. [cited 2021 Feb 9] Available from: https://doi.org/https://doi.org/10.1016/j.biopsych.2020.01.016

58. Thalhauser CJ, Komarova NL. Alzheimer's disease: rapid and slow progression. J R Soc Interface. 2012;9:119–26 [cited 2021 Feb 9] Available from: /pmc/articles/PMC3223623/.

59. Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D. Early diagnosis of Alzheimer's disease with deep learning. 2014 IEEE 11th Int Symp Biomed Imaging, ISBI 2014. Institute of Electrical and Electronics Engineers Inc.; 2014. p. 1015–1018.

60. Busche MA, Hyman BT. Synergy between amyloid-β and tau in Alzheimer's disease. Nat. Neurosci. 2020:1183–93 [cited 2021 May 26] Available from: https://pubmed.ncbi.nlm.nih.gov/32778792/.

61. Bennett RE, DeVos SL, Dujardin S, Corjuc B, Gor R, Gonzalez J, et al. Enhanced tau aggregation in the presence of amyloid β. Am J Pathol. 2017;187:1601–12. [cited 2021 Jul 21] Available from: https://pubmed.ncbi.nlm.nih.gov/28500862/

62. Götz J, Chen F, van Dorpe J, Nitsch RM. Formation of Neurofibrillary Tangles in P301L Tau Transgenic Mice Induced by $A\beta_{42}$ Fibrils. Science. 2001;293(5534):1491-5. https://doi.org/10.1126/science.1062097.

63. Hanseeuw BJ, Betensky RA, Jacobs HIL, Schultz AP, Sepulcre J, Becker JA, et al. Association of Amyloid and Tau With Cognition in Preclinical Alzheimer Disease. JAMA Neurol. 2019;76(8):915. https://doi.org/10.1001/jamaneurol.2019.1424.

Sadiq *et al. Alzheimer's Research & Therapy*   (2022) 14:16

Page 14 of 14

64. Sperling RA, Mormino EC, Schultz AP, Betensky RA, Papp K V., Amariglio RE, et al. The impact of amyloid-beta and tau on prospective cognitive decline in older individuals. Ann Neurol. 2019;85:181–93. [cited 2021 Jul 21] Available from: https://pubmed.ncbi.nlm.nih.gov/30549303/

65. Adams JN, Maass A, Harrison TM, Baker SL, Jagust WJ. Cortical tau deposition follows patterns of entorhinal functional connectivity in aging. Elife. 2019;8 [cited 2021 Nov 5] Available from: https://pubmed.ncbi.nlm.nih.gov/31475904/.

66. Fortea J, Vilaplana E, Alcolea D, Carmona-Iragui M, Sánchez-Saudinos MB, Sala I, et al. Cerebrospinal fluid β-amyloid and phospho-tau biomarker interactions affecting brain structure in preclinical Alzheimer disease. Ann Neurol. 2014;76:223–30 [cited 2021 Nov 5] Available from: https://pubmed.ncbi.nlm.nih.gov/24852682/.

67. Jacobs HIL, Hedden T, Schultz AP, Sepulcre J, Perea RD, Amariglio RE, et al. Structural tract alterations predict downstream tau accumulation in amyloid-positive older individuals. Nat Neurosci. 2018;21:424–31. [cited 2021 Jul 21] Available from: https://pubmed.ncbi.nlm.nih.gov/29403032/

68. Huang Y, Xu J, Zhou Y, Tong T, Zhuang X. Diagnosis of Alzheimer's disease via multi-modality 3D convolutional neural network. Front Neurosci. Frontiers; 2019;13:509.

69. Qiu S, Joshi PS, Miller MI, Xue C, Zhou X, Karjadi C, et al. Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. Brain. 2020;143:1920–33. [cited 2021 Jul 21] Available from: https://academic.oup.com/brain/article/143/6/1920/5827821

70. Dujardin S, Commins C, Lathuiliere A, Beerepoot P, Fernandes AR, Kamath TV, et al. Tau molecular diversity contributes to clinical heterogeneity in Alzheimer's disease. Nat Med. 2020;26:1256–63 [cited 2021 Mar 17] Available from: https://doi.org/10.1038/s41591-020-0938-9.

71. Blom ES, Giedraitis V, Zetterberg H, Fukumoto H, Blennow K, Hyman BT, et al. Rapid progression from mild cognitive impairment to alzheimer's disease in subjects with elevated levels of tau in cerebrospinal fluid and the Apoe ε4/ε4 genotype. Dement Geriatr Cogn Disord. Dement Geriatr Cogn Disord; 2009;27:458–64. [cited 2021 Mar 17] Available from: https://pubmed.ncbi.nlm.nih.gov/19420940/

72. Buchhave P, Minthon L, Zetterberg H, Wallin ÅK, Blennow K, Hansson O. Cerebrospinal fluid levels of β-amyloid 1-42, but not of tau, are fully changed already 5 to 10 years before the onset of Alzheimer dementia. Arch Gen Psychiatry. 2012;69:98–106. [cited 2021 Jul 15] Available from: https://jamanetwork.com/journals/jamapsychiatry/fullarticle/1107443

73. Li R, Zhang W, Il SH, Wang L, Li J, Shen D, et al. Deep learning based imaging data completion for improved brain disease diagnosis. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). Springer Verlag. 2014:305–12 [cited 2021 Apr 5] Available from: https://pubmed.ncbi.nlm.nih.gov/25320813/.

74. Li H, Habes M, Wolk DA, Fan Y. A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. Alzheimer's Dement. 2019;15:1059–70 [cited 2021 Feb 9] Available from: https://pubmed.ncbi.nlm.nih.gov/31201098/.

Publisher's Note