Article

Subtyping of Alzheimer's Disease Using Multimodal Transformer

Diego Machado-Reyes [1] , Hanqing Chao [1] , Juergen Hahn [1], , Li Shen [2], , and Pingkun Yan [1,*] ,
for the Alzheimer's Disease Neuroimaging Initiative

[1]   Department of Biomedical Engineering and the Center for Biotechnology and Interdisciplinary Studies, 110 8th St, Troy, 12180, New York, USA; machad@rpi.edu(D.M.R.); chaoh.rpi@gmail.com (H.C.); hahnj@rpi.edu (J.H.); yanp2@rpi.edu (P.Y.)

[2]   Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, 19104, Pennsylvania, USA; li.shen@pennmedicine.upenn.edu

*   Correspondence: yanp2@rpi.edu

Abstract: Alzheimer's disease (AD) is the most prevalent neurodegenerative disease, yet its current treatments are limited to stopping disease progression. Moreover, the effectiveness of these treatments remains uncertain due to the heterogeneity of the disease. Therefore, it is essential to identify disease subtypes at a very early stage. Current data-driven approaches can classify subtypes during later stages of AD, but face challenges when predicting in the asymptomatic or prodromal stage. Furthermore, most existing models lack explainability in their classification and rely solely on a single modality for assessment, limiting the scope of their analysis. Thus, we propose a multimodal framework that utilizes early-stage indicators, including imaging, genetics, and clinical assessments, to classify AD patients into subtypes at an early stage. In our framework, we introduce a tri-modal co-attention mechanism (Tri-COAT) to explicitly capture cross-modal feature associations. **[DM: Data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Slow progressing = 177, Intermediate = 302, Fast = 15) was used to train and evaluate Tri-COAT using a 10-fold stratified cross-testing approach.]** Our proposed model outperforms baseline models and sheds light on essential associations across multimodal features supported by the known biological mechanisms. **[DM: The multimodal design behind Tri-COAT allowed it to achieve the highest performance, while simultaneously provide interpretability to the model predictions through the co-attention mechanism.]**

Keywords: Disease subtyping; artificial intelligence; multimodal biomarker; transformer network; Alzheimer's disease

1. Introduction

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder, which affects over 6.5 million people in the US alone and its rate is expected to keep increasing [1]. Current therapies for AD are mainly focused on the management of symptoms and promising drugs that can slow the progression of the disease [2–4]. Therefore, early diagnosis of neurodegenerative diseases is crucial. However, early diagnosis of AD presents a significant challenge since **[DM: memory loss only develops in the MCI stage of the AD continuum and cognitive decline]** can vary among patients due to the disease's heterogeneity [5]. Therefore, it is crucial to develop methods capable of characterizing the factors that influence disease progression and identifying specific subtypes for individual patients at an early stage.

AD is traditionally diagnosed based on characteristic cognitive decline and behavioral deficits that do not become apparent until intermediate to late stages of the disease. Early stage indicators such as imaging-based and fluid biomarkers have shown great potential for early detection of AD [6]. Similarly, recent approaches have continued to show the promising results from connectomes for early-diagnosis [7]. **[DM: Blood and CSF biomarkers are now standard methods to diagnose early AD patients [8–10] showing great potential for subtyping AD patients [11]. AD subtypes have been previously identified based on hallmark AD biomarkers obtained from brain imaging such as beta-amyloid [12] and tau accumulation [13,14].]** Data driven approaches have focused on classifying patients into subtypes based on the disease progression and mild cognitive impairment (MCI, a prodromal stage of AD) to

Version March 30, 2024 submitted to *J. Pers. Med.*

https://www.mdpi.com/journal/jpm

AD conversion [15–19]. Current methods for subtyping of AD and related disorders have focused mainly on using longitudinal data from clinical assessments for unsupervised learning [20–22]. For clustering approaches such as in [15,17,20,22], the authors subtype AD patients using single modalities at baseline, such as blood markers [17], genomic data [20], or traits derived from imaging from longitudinal measurements [22]. These single modality approaches at baseline data have shown the potential of early stage indicators for AD subtyping. However, most of them fail to show how they relate biologically to AD development or use multiple time points, which hinders the ability to diagnose patients at early stages after AD onset.

Deep learning models have effectively identified diagnostic groups [23] and subtypes [18] using multimodal imaging data and correlation-based approaches that allow a greater explainability of of the relationships between the features learned by the model. Nevertheless, these are limited by two key factors, using only imaging data, and for correlation-based approaches, these treat the clustering goal indirectly. Works on related disorders have shown the relevance of multimodal approaches [24] using non-negative matrix factorization and Gaussian mixture models, or employing autoencoders and long short term memory (LSTM) networks to learn deep embeddings of disease progression [25]. However, this requires longitudinal data that involve several time points. While several clustering and deep learning-based approaches have high accuracy when having multiple time points, the performance decreases significantly given only baseline data. This is driven by the subtle expression of the symptoms tracked in the clinical assessments limiting the scope of the models. Early stage indicators such as imaging traits or genomic risk factors are rarely used, and are simply aggregated to clinical assessments as additional indicators. Therefore, it is essential to target early stage indicators such as genetics, imaging and cognitive assessments.

Multimodal deep learning approaches can combine different modalities and provide a much more informed picture of disease drivers and aid in the disease subtyping [26]. However, it is not a trivial task to identify the relevant features across modalities and how to fuse them. The rest of this section first reviews the related works on multimodal fusion and then provides an overview of our proposed method.

1.1. Multimodal Fusion

Multimodal fusion, while very promising, poses new challenges. There are multiple ways to fuse the data and stages of encoding where to fuse. The effectiveness of the strategy varies depending on data modalities and tasks. One of the key factors is the similarity between the modalities. Highly heterogeneous data, such as imaging, genetics, and clinical data, might not be immediately fused. Their difference in type, signal-to-noise ratio, and dimensionality makes it very challenging to combine them without first projecting them into a similar space. The relationship between input and output is equally important to consider when fusing data. For example, clinical assessments reflect the direct impacts of the disease, while genetic data describes the building blocks of cells. The phenotype-related information available in clinical assessments requires considerably less processing than what genomic data might require to find the connection with the disease.

As illustrated in Fig. 1, the existing multimodal fusion strategies can be grouped into three main categories [27], including early, intermediate (also called joint), and late fusion. It is essential to choose the right approach based on the task at hand and the data used. Early stage fusion has shown very promising results in recent vision-language models [28,29], while late-fusion strategies have traditionally been very effective in aggregating machine learning models decisions. Early and late stage fusion strategies, while effective for certain tasks, are not ideal for AD subtyping using multimodal data. Early stage fusion struggles with dealing with highly heterogeneous and differently biologically-related data such as genetics, imaging and clinical. These require further encoding to first learn highly informative representations for every modality and condense them into a similar latent space. While late fusion strategies can be very effective at aggregating the decisions based on each modality, they cannot learn the feature relationships across modalities severely limiting the usefulness of the model. The intermediate fusion approach tackles both challenges by first learning the crucial patterns associated with each modality. In the next stage, it uses the condensed patterns from each modality to learn the cross-modal relationships. This enables a more harmonious fusion of the heterogeneous modalities.
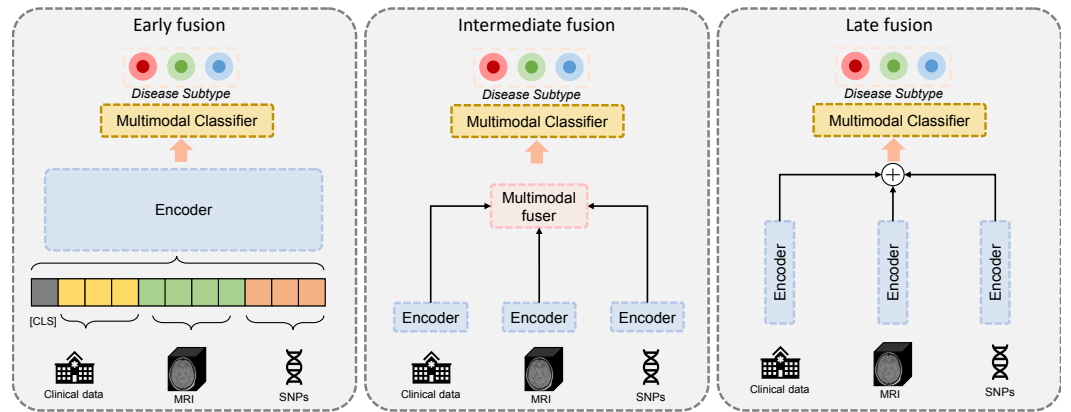
Figure 1. The three main multimodal fusion strategies, early, intermediate and late fusion, for deep learning methods.

1.2. Overview and Contributions

Despite the potential advantages of multimodal approaches, several technical challenges exist to effectively leverage the key data of each modality. The high heterogeneity of the data modalities for AD subtyping, and the explicit learning of the cross-modal interactions need to be addressed. Previous approaches in related fields have proposed dual co-attention mechanisms to explicitly learn the cross-modal feature interaction and joint data representations [30]. While they have shown very promising results, these have yet to be explored for the subtyping of neurodegenerative disease and its corresponding modalities. Moreover, in AD and related disorders subtyping, **[DM: three critical but highly heterogeneous modalities are needed to be fused]** , i.e., imaging, genetics and clinical data. Therefore, in this paper, a Tri-modal co-attention (Tri-COAT) framework **[DM: is proposed]**, that can explicitly learn the interactions between the multimodal features towards the task of classifying subtypes. While deep learning models for disease assessment promise improved accuracy, they remain limited in the interpretability of their results. This is a major entry barrier for the inclusion of deep learning models in the medical field.

Our contributions in this paper are two fold in both the application and technique. **[DM: First,]** our framework incorporates features of three early stage biomarker modalities and provides a cutting-edge approach to the subtyping of early neurodegenerative diseases. **[DM: Second, regarding the technical innovation,]** our new Tri-modal co-attention can efficiently and explicitly learn the interactions between highly heterogeneous modalities, encode the information into a joint representation, and provide explainability to the cross-modal interactions. The proposed Tri-COAT achieved state-of-the-art performance on the landmark dataset Alzheimer's Disease Neuroimaging Initiative (ADNI)[31], and provided key insights into the biological pathways leading to neurodegenerative disease development.

[DM: The rest of the paper is organized as follows. In section 2, the methods are described breaking down each component of Tri-COAT. Next, in section 3, the dataset and experimental design are delineated. In section 4, the results are presented and discussed. Finally, in section 5, conclusions are drawn and future directions are presented.]

2. Method

Our proposed framework can be divided into two main parts. As seen in Fig. 2, single modality encoders are first built using transformer modules to learn feature representations for each modality. Second, the Tri-COAT mechanism explicitly learns the critical cross-modal feature relationships and uses that to weigh the feature representation. The jointly learned representation goes through a multilayer perceptron (MLP) for disease subtype classification.

2.1. Single-modality encoding

Three branches encode each modality individually. Each branch is comprised of a transformer encoder with several transformer layers. **[DM: This is inspired by previous works in which transformer**
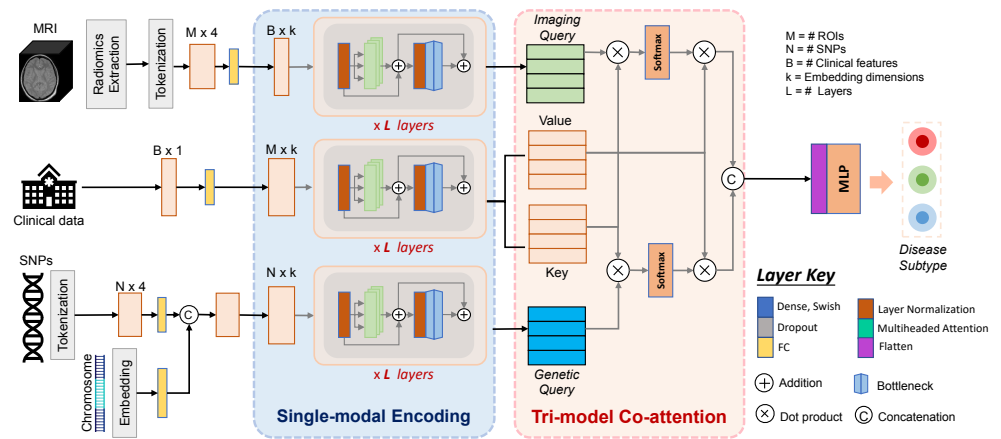
Figure 2. Illustration of the proposed framework for AD subtyping, consisting of two main sections: (a) single modality encoding and (b) tri-modal attention and joint encoding.

models have been proposed for imaging derived connectomes [32] and genotype data [33]]. Each branch learns representations of a modality to later combine them into a joint representation through the Tri-COAT mechanism.

Imaging modality. The imaging feature encoder branch uses MRI derived quantitative traits as input. These quantitative traits are derived from T1 weighted MRI scans. The scans are first segmented based on the FreeSurfer atlas for cross-sectional processing. Next, for each reconstructed regions of interest (ROI) the cortical region, cortical volume, thickness average, thickness standard deviation, and surface area are calculated. Further details are described in section 3.1. The imaging traits are then used to build tokens, where each token represents an ROI in the brain and is comprised of four imaging derived traits (Cortical Thickness Average, Cortical Thickness Standard Deviation, Surface Area, Volume from Cortical Parcellation). Let $X_I \in \mathbb{R}^{M \times 4}$ represent the imaging input to the proposed model, where $M = 72$ and is the number of ROIs. After, the token dimensions are expanded through a fully-connected layer to match the model dimensions k. The imaging tokenization allows to build an initial representation for each ROI rather than each trait, leading to a smaller number of input features, and a more biologically informative and interpretable input.

Genetic modality. The genotype branch has as input single nucleotide polymorphism (SNP) data. After quality control and preprocessing of the genotype data as described in section 3.1, tokens for each SNP are built. Each token is comprised by the allele dosage from the patient, the corresponding odds ratio and rare allele frequency obtained from the most recent AD GWAS study, and whether the SNP is within an intergenic region (regulatory region) as a binary label. Then, the token dimensions are expanded through a fully-connected layer to $k/2$. Let $X_{SNP} \in \mathbb{R}^{N \times k/2}$ represent the genotype input to the model, where $N = 70$ and is the number of SNPs filtered out (see section 3.1 for details). Moreover, based on the chromosome for which each SNP is located in an additional embedding for each SNP can be built. By including the chromosome embedding, location knowledge for each SNP can be incorporated. Using an embedding layer, an embedding for each chromosome can be obtained $X_{Chr} \in \mathbb{R}^{N \times k/2}$. Finally, X_{SNP} and X_{Chr} are concatenated to obtain the final genotype embedding $X_G \in \mathbb{R}^{N \times k}$. Similarly to the imaging data encoding, the genetic tokenization allows to build more informative input structures to the genetic encoder. By providing additional attributes for each SNP beyond the patient mutation status, the model can learn richer patterns of characteristics that relate each SNP.

Clinical modality. The clinical data is already very closely related to the outcome of interest and contains only few features; therefore, no further extensive tokenization is performed. As there is just one value per clinical assessment, the tokens are directly built with one dimension. Let $X_C \in \mathbb{R}^{B \times 1}$ represent the clinical input to our model, where $B = 7$ and is the number of clinical features. Next, the token dimensions are expanded through a fully-connected layer to match the model dimensions k.
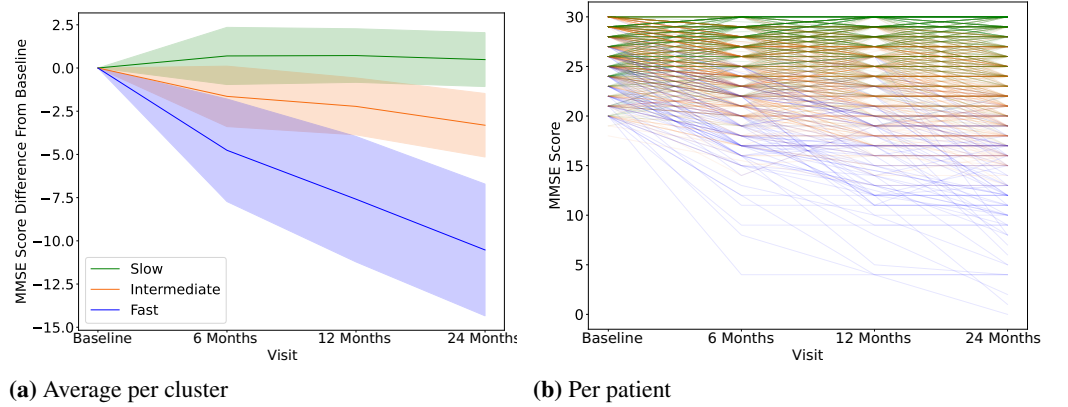
(a) Average per cluster **(b)** Per patient

Figure 3. AD subtype clusters based on the decrease of MMSE at each visit. (a) each line represents the average score across patients for each cluster and the shadow represents one standard deviation. (b) individual lines per patient are plotted.

Single modality encoders. After the tokenization of each modality, they are fed into independent transformer encoders with L layers. The full process of the L-th layer in our transformer encoder is formulated as:

$$F_l' = \text{MHA}(\text{LN}(F_{l-1})) + F_{l-1}, \tag{1}$$

and

$$F_l = \text{FF}(\text{LN}(F_l')) + F_l', \tag{2}$$

where $\text{LN}(\cdot)$ is the normalization layer [34], and $\text{MHA}(\cdot)$ is multihead self-attention [35].

2.2. Tri-modal co-attention

After each transformer encoder has learned a new representation for each modality, these are then used to learn the cross-modal feature relationships to guide the co-attention process on the clinical branch. In other words, the imaging and genomic features are employed to modulate the clinical learning process by highlighting the key hidden features that share relationships across modalities. The intuition behind the proposed approach is that as the clinical data is most closely related to the disease phenotype, this branch will carry most of the necessary information to classify the patients. Nevertheless, the imaging and genomic data provide also valuable information. The idea is analogous to the clinical data being the subject and verb in a sentence while the imaging and genomic data are the adjectives and adverbs. These two elements enrich the representation of the patient health status, analogous to enriching a sentence for a fuller meaning.

Let $X_{Emb} \in \mathbb{R}^{\{M,N,B\} \times k}$ represent the learned representation of a given single-modality encoder. These become then query matrices for the genetics Q_G and imaging Q_I data, and key K_C, value V_C matrices for the clinical data. Following an attention mechanism structure, the co-attention between two modalities is computed as:

$$\text{CoAttn}(\{G, I\}, C) = \text{softmax}\left(\frac{Q_{\{G,I\}} K_C^T}{\sqrt{d_k}}\right) V_C^T. \tag{3}$$

Next, the resulting co-attention filtered value matrices are then concatenated to obtain a final joint representation. This joint representation is then flattened and used to classify the patients into the clusters through an MLP.

3. Materials and Experiments

3.1. Dataset

The Alzheimer's Disease Neuroimaging Initiative (ADNI)[31] database is a landmark dataset for the advancement of our understanding of Alzheimer's disease. ADNI[31] is comprised by a wide

Table 1. Subject data distribution reported as mean $\pm$ standard deviation for MMSE and age, and counts per category for number of participants and gender.

	Slow	**Intermediate**	**Fast**
Participants	177	302	15
MMSE (Baseline)	27.35 ± 2.51	27.66 ± 1.86	24.93 ± 3.55
MMSE (24 months)	28.15 ± 2.15	23.86 ± 3.68	15.9 ± 4.84
Age	73.26 ± 7.82	72.44 ± 7.55	71.22 ± 3.922
Gender	M: 102 F: 75	M: 185 F: 117	M: 9 F: 6

range of data modalities including MRI and PET images, genetics, cognitive tests, CSF and blood biomarkers (for up-to-date information, see www.adni-info.org). Longitudinal data for all subjects was selected as up to two years of progression after disease onset due to very high missingness rate (percentage of data points missing across patients for a given time point) present for time points posterior to the two years. Subjects were clustered using k-means in 3 main groups based on their Mini-Mental State Examination (MMSE) scores as can be seen in Fig. 3. These clusters match the cognitive decline rate of patients over time. The MMSE score at each visit (baseline, 6 months, 12 months, and 24 months) was used to determine the cognitive decline for each patient. As each patient may have a different starting level at baseline, the baseline measurement is subtracted from each of the following time points; thus all patients start at 0. Then, using k-means clustering, using $k = 3$, slow, intermediate and fast cognitive decline groups are defined. As seen in Table 1 the raw average MMSE score at baseline is comparable across all groups with a steep decrease for the fast and intermediate groups at 24 months. The slow or otherwise stable group MMSE score at 24 months is comparable to the one at the initial stage. On the other hand, all 3 groups are age matched. Similarly, gender distributions across groups is maintained with male subjects representing approximately 60% of the subjects for each group.

Data preprocessing: The data for the imaging and genotype modalities was processed previously to the tokenization process following best practices for the corresponding modality as described below.

- **Imaging**: The FreeSurfer image analysis suite (http://surfer.nmr.mgh.harvard.edu/) is used to conduct cortical reconstruction and volumetric segmentation. T1 weighted MRI scans are segmented based on the FreeSurfer atlas for cross-sectional processing, enabling group comparison at a specific time point [36]. For each reconstructed cortical region, cortical volume, thickness average, thickness standard deviation, and surface area measurements are labeled by the 2010 Desikan-Killany atlas. The UCSF ADNI team conducted this process [37], and more information can be found at http://adni.loni.usc.edu.
- **Genotype**: The genotype variants were filtered using the intersection between the *List of AD Loci and Genes with Genetic Evidence Compiled by ADSP Gene Verification Committee* and the most recent genome wide association study (GWAS) on AD [38]. The odds Ratios, rare allele frequency and intergenic region binary trait were obtained from the most recent GWAS study with accession number (GCST90027158), accessed through the GWAS catalog [39]. Furthermore, the genotype variants were processed for sample and variant quality controls using PLINK1.9 [40].
- **Clinical**: The clinical assessment features correspond to seven different cognitive metrics available through ADNI [31]. These are Logical Memory - Delayed Recall (LDELTOTAL), Digit Symbol Substitution (DIGITSCOR), Trail Making Test B (TRABSCOR) and Rey Auditory Verbal Learning Test (RAVLT) scores: immediate, learning, forgetting and percent forgetting. Values for imaging and clinical modalities were normalized in each train set before they were used as inputs to the network.

3.2. Experimental design

All models underwent training and evaluation using a 10-fold stratified cross-testing approach. Initially, the entire dataset was divided into 10 folds, with one fold reserved for testing and the

Table 2. Mean AUROC $\pm$ SD of 10-fold cross-testing results. The proposed model significantly outperformed all the baseline models. The statistical significance was evaluated by paired *t*-test with $\alpha = 0.005$, except for the entry where $\alpha = 0.05$.

Method	Full	Imaging	Genetics	Clinical
SVM	*0.705 $\pm$ 0.036*	0.669 $\pm$ 0.060	0.525 $\pm$ 0.034	0.639 $\pm$ 0.078
RF	0.684 $\pm$ 0.048	0.677 $\pm$ 0.052	0.505 $\pm$ 0.031	0.659 $\pm$ 0.087
Stage-wise fusion	0.641 $\pm$ 0.017	0.557 $\pm$ 0.096	0.562 $\pm$ 0.078	0.655 $\pm$ 0.057
Tri-COAT	**0.734 $\pm$ 0.076**	0.648 $\pm$ 0.056	0.539 $\pm$ 0.084	0.697 $\pm$ 0.063

remaining nine for training. Subsequently, this training set was further divided into 5 folds for a 5-fold stratified cross-validation process for hyperparameter tuning. This robust framework was designed to prevent any data leakage. The optimal hyperparameters were determined during each experimental run by selecting the best-performing model based on the validation set. Each of the 10 test sets was evaluated 5 times, using hyperparameters determined by the validation sets, resulting in a total of 50 evaluations for each method. Predictions were evaluated using the area under the receiver operating curve (AUROC). A one-vs-one strategy is employed where the average AUROC of all possible pairwise combinations of classes is computed for a balanced metric. This is implemented using Sci-Kit Learn API [41], which implements the method described in [42]. The mean AUROC and standard deviation across all 50 runs are reported for each method in Table 2. The model was compared against the stage-wise deep learning intermediate fusion model introduced in [43] and several well-established traditional ML models - random forest (RF), support vector machine (SVM) with radial-basis function (RBF) kernel. Similarly, each of the branches of the model was used as comparison using a series of transformer encoder layers and MLP head for classification.

Tri-COAT consist of four transformer layers for each of the single modality encoders, with four attention heads per transformer layer. The tri-model co-attention process is done in a single-head attention mechanism. The classifying MLP has one hidden layer with 256 units. Embedding dimension of $k = 256$ was used for all modalities. The model dimensions for the single-modality encoders were kept at 256 and all throughout as this combination achieved the best results on the validation set. The final MLP had the concatenated class tokens resulting from the tri-modal co-attention module and computed the output logits for each one of the three possible classes. Adam was used to optimize both Tri-COAT and the stage-wise MLP model using learning rates of 0.0001 for Tri-COAT and 0.0001 for the stage-wise fusion benchmarking model. All deep learning models were trained using cross-entropy loss. All deep learning models were implemented using PyTorch, while the RF and SVM models were implemented using scikit-learn. All deep learning models were trained for 100 epochs and the best checkpoint, meaning epoch with the highest AUROC on the validation set, was selected for model evaluation. The stage-wise deep learning fusion model had dimensions of 64 units for the single-modality layers, 32 for the second-stage and 16 for the final stage. The model dimensions were selected following the described hyperparameters in [43]. The SVM used an RBF kernel, regularization parameter C of 1. The random forest used gini as its criterion for leaf splitting, 100 trees and no maximum depth.

4. Results and Discussion

4.1. Clustering, label definition

Following the literature, the number of clusters was set to three main groups [44–46]. The MMSE score was used as an indicator of mental decline. Based on the speed of the progression of the mental decline over a period of two years, three groups were defined, i.e., fast, intermediate and slow progressing subtypes. K-means clustering was used to assign each subject to one of the groups. Through this process labels were defined for all subjects. Only baseline data was used as input to Tri-COAT and all competing models. Based on the baseline data Tri-COAT was able to effectively classify the subjects into their corresponding subtypes.
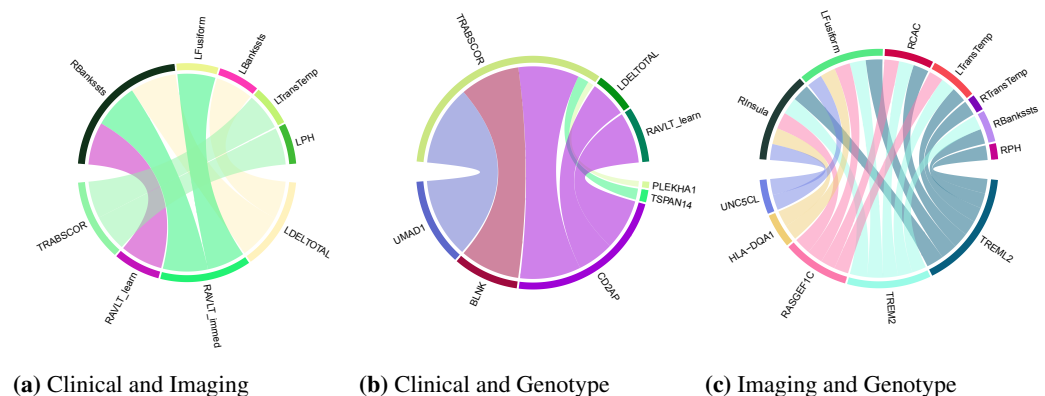
(a) Clinical and Imaging   (b) Clinical and Genotype   (c) Imaging and Genotype

Figure 4. Cross modal associations of AD key biomarkers visualized from the learned co-attention.

4.2. AD subtype classification

As seen in Table 2, Tri-COAT outperformed the single modality ablations and baseline models, achieving an average AUROC of 0.733 ± 0.070 across all test sets in the 10-fold cross-testing framework. For the single-modality ablation studies, each modality was used independently to classify the AD subtype. For Tri-COAT, a single-modality transformer encoder backbone and MLP head were used. For the stage-wise fusion model, it was adapted to MLPs using the same number of hidden layers as the first plus last stage of the multimodal version. For SVM and RF, no variations were required. Each modality was evaluated using the same 10-fold cross testing - 5 fold-cross validation hyperparameter tuning framework. Moreover, the single modality ablation models outperformed their baseline counterparts for the clinical and genetics modalities. In contrast to this the baselines achieved better performances for the imaging derived traits. For the three modalities, clinical achieves the best classification AUROC, followed by imaging and genetics. This is expected as biologically, the same order follows for the closest relation between the observed phenotype and the mechanisms behind it. Clinical (cognitive) assessments are the closest to the MMSE metric, followed by imaging (changes in the brain morphology) which is directly related to the observed phenotype and genetics being the farthest apart from the expressed symptoms. Both comparative models and Tri-COAT achieve higher performances in their multimodal configuration compared to single modalities, agreeing with previous literature regarding multimodal approaches for classification of AD and related disorders.

Furthermore, as seen in Table 3, Tri-COAT outperformed variations of itself using alternative fusion strategies. The early fusion model considerably underperforms achieving an AUROC of 0.571 $\pm$ 0.053 because of its limited capabilities to simultaneous encode highly heterogeneous data with distinct biological level relationships to the outcome. Similarly, the late fusion model underperforms as it is limited to joining the predictions from each branch and cannot learn the relationships between the different modalities in the latent space.

4.3. Biomarker Associations Learned by Co-attention

One of the key advantages of Tri-COAT compared to the baseline models and traditional deep learning approaches is the ability to learn insights into the cross-modal feature associations. In order to explore the learned relationships, the model with highest test AUROC from the evaluation framework was selected for attention visualization in which the learned attention scores were averaged across all test subjects. Chord plots were drawn using the circlize R library [47] to visualize the cross-modal attention. As seen in Fig. 4, Tri-COAT identified key associations between the Trails making test B score (TRABSCORE) in clinical-imaging and clinical-genetics. This score tests for cognitive ability of the patient for working memory and secondarily task-switching ability [48]. Clinical literature shows strong correlation between gyri structures - temporal gyrus and Parahippocampal gyrus (LTransTemp, LPH) and TRABSCORE [49]. Similarly, for the clinical-genotype association, TRABSCORE was found to be associated with the CD2AP gene which has

Table 3. Mean AUROC $\pm$ SD of 10-fold cross-testing results. The proposed model significantly outperformed all the baseline models. The statistical significance was evaluated by paired *t*-test with $\alpha = 0.005$, except for the entry where $\alpha = 0.05$.

Method	AUROC
Early	*0.571 $\pm$ 0.053*
Late	0.604 $\pm$ 0.048
Tri-COAT	**0.734 $\pm$ 0.076**

been clinically identified as a driver for the AD hallmark - neurofibrillary tangles (NFT) in the temporal gyrus region [50]. This is a very exciting finding for our network as it establishes a putative relationship between genetics (CD2AP gene), brain ROIs (temporal gyrus) and clinical symptoms (TRABSCORE).

5. Conclusions

AD is the most prevalent neurodegenerative disease, and all current treatments are limited to slowing disease progression. Therefore, early diagnosis is essential. Furthermore, there are multiple subtypes with different rates of cognitive decline. In order to move closer to personalized medicine, it is essential to have a better understanding of the heterogeneity surrounding the disease development. However, early subtyping is a very challenging task. Our proposed model was able to effectively classify AD patients into three main subtypes using prodromal factors measured at baseline.

Moreover, the model was able to identify multiple putative cross-modal biomarker networks. [DM: The putative biomarkers provide enhanced interpretability to Tri-COAT and shed light into possible exciting therapeutic targets.] Nevertheless, the generalizability of applying the learned features to other datasets remains to be tested.

[DM: The future directions are very exciting as Tri-COAT could be extended] to other heterogeneous neurodegenerative diseases such as Parkinson's disease. [DM: Moreover, as shown in this work, multimodal approaches achieved the best results. A promising future direction is to incorporate further modalities such as PET imaging and transcriptomic data. PET imaging could provide further clarity towards the accumulation of fluid biomarkers and their impact towards neurodegeneration. Similarly, transcriptomic data could provide an intermediate biological step between the genotype and brain endophenotypes. These could lead to enhanced understanding of the underlying mechanisms and provide further therapeutic targets.]

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not applicable

Data Availability Statement: All data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

References

1. 2022 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **2022**, *18*, 700–789. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/alz.12638, https://doi.org/10.1002/alz.12638.

2. Dhillon, S. Aducanumab: First Approval. *Drugs* **2021**, *81*, 1437–1443. https://doi.org/10.1007/s40265-021-01569-z.

3. Swanson, C.J.; Zhang, Y.; Dhadda, S.; Wang, J.; Kaplow, J.; Lai, R.Y.K.; Lannfelt, L.; Bradley, H.; Rabe, M.; Koyama, A.; et al. A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-AB protofibril antibody. *Alzheimer's Research & Therapy* **2021**, *13*, 80. https://doi.org/10.1186/s13195-021-00813-8.

4. Shcherbinin, S.; Evans, C.D.; Lu, M.; Andersen, S.W.; Pontecorvo, M.J.; Willis, B.A.; Gueorguieva, I.; Hauck, P.M.; Brooks, D.A.; Mintun, M.A.; et al. Association of Amyloid Reduction After Donanemab Treatment With Tau Pathology and Clinical Outcomes: The TRAILBLAZER-ALZ Randomized Clinical Trial. *JAMA Neurology* **2022**, *79*, 1015–1024. https://doi.org/10.1001/jamaneurol.2022.2793.

5. Foster, N.L.; Bondi, M.W.; Das, R.; Foss, M.; Hershey, L.A.; Koh, S.; Logan, R.; Poole, C.; Shega, J.W.; Sood, A.; et al. Quality improvement in neurology. *Neurology* **2019**, *93*, 705–713. Publisher: Wolters Kluwer, https://doi.org/10.1212/WNL.0000000000008259.

6. Hassen, S.B.; Neji, M.; Hussain, Z.; Hussain, A.; Alimi, A.M.; Frikha, M. Deep learning methods for early detection of Alzheimer's disease using structural MR images: a survey. *Neurocomputing* **2024**, *576*, 127325. https://doi.org/10.1016/j.neucom.2024.127325.

7. Zhang, S.; Zhao, H.; Wang, W.; Wang, Z.; Luo, X.; Hramov, A.; Kurths, J. Edge-centric effective connection network based on muti-modal MRI for the diagnosis of Alzheimer's disease. *Neurocomputing* **2023**, *552*, 126512. https://doi.org/10.1016/j.neucom.2023.126512.

8. McKhann, G.M.; Knopman, D.S.; Chertkow, H.; Hyman, B.T.; Jack Jr., C.R.; Kawas, C.H.; Klunk, W.E.; Koroshetz, W.J.; Manly, J.J.; Mayeux, R.; et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia* **2011**, *7*, 263–269. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1016/j.jalz.2011.03.005, https://doi.org/10.1016/j.jalz.2011.03.005.

9. Dubois, B.; Feldman, H.H.; Jacova, C.; Hampel, H.; Molinuevo, J.L.; Blennow, K.; DeKosky, S.T.; Gauthier, S.; Selkoe, D.; Bateman, R.; et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet Neurology* **2014**, *13*, 614–629. Publisher: Elsevier, https://doi.org/10.1016/S1474-4422(14)70090-0.

10. Tao, Q.Q.; Lin, R.R.; Wu, Z.Y. Early Diagnosis of Alzheimer's Disease: Moving Toward a Blood-Based Biomarkers Era. *Clinical Interventions in Aging* **2023**, *18*, 353–358. https://doi.org/10.2147/CIA.S394821.

11. Dubois, B.; von Arnim, C.A.F.; Burnie, N.; Bozeat, S.; Cummings, J. Biomarkers in Alzheimer's disease: role in early and differential diagnosis and recognition of atypical variants. *Alzheimer's Research & Therapy* **2023**, *15*, 175. https://doi.org/10.1186/s13195-023-01314-6.

12. Collij, L.E.; Salvadó, G.; Wottschel, V.; Mastenbroek, S.E.; Schoenmakers, P.; Heeman, F.; Aksman, L.; Wink, A.M.; Berckel, B.N.; van de Flier, W.M.; et al. Spatial-Temporal Patterns of B-Amyloid Accumulation. *Neurology* **2022**, *98*, e1692–e1703. Publisher: Wolters Kluwer, https://doi.org/10.1212/WNL.0000000000200148.

13. Bejanin, A.; Schonhaut, D.R.; La Joie, R.; Kramer, J.H.; Baker, S.L.; Sosa, N.; Ayakta, N.; Cantwell, A.; Janabi, M.; Lauriola, M.; et al. Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. *Brain* **2017**, *140*, 3286–3300. https://doi.org/10.1093/brain/awx243.

14. Vogel, J.W.; Young, A.L.; Oxtoby, N.P.; Smith, R.; Ossenkoppele, R.; Strandberg, O.T.; La Joie, R.; Aksman, L.M.; Grothe, M.J.; Iturria-Medina, Y.; et al. Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nature Medicine* **2021**, *27*, 871–881. Publisher: Nature Publishing Group, https://doi.org/10.1038/s41591-021-01309-6.

15. Mitelpunkt, A.; Galili, T.; Kozlovski, T.; Bregman, N.; Shachar, N.; Markus-Kalish, M.; Benjamini, Y. Novel Alzheimer's disease subtypes identified using a data and knowledge driven strategy. *Scientific Reports* **2020**, *10*, 1327. Number: 1 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41598-020-57785-2.

16. Badhwar, A.; McFall, G.P.; Sapkota, S.; Black, S.E.; Chertkow, H.; Duchesne, S.; Masellis, M.; Li, L.; Dixon, R.A.; Bellec, P. A multiomics approach to heterogeneity in Alzheimer's disease: focused review and roadmap. *Brain* **2020**, *143*, 1315–1331. https://doi.org/10.1093/brain/awz384.

17. Martí-Juan, G.; Sanroma, G.; Piella, G.; Consortium, f.t.A.D.N.I.a.t.A.D.M. Revealing heterogeneity of brain imaging phenotypes in Alzheimer's disease based on unsupervised clustering of blood marker profiles. *PLOS ONE* **2019**, *14*, e0211121. Publisher: Public Library of Science, https://doi.org/10.1371/journal.pone.0211121.

18. Feng, Y.; Kim, M.; Yao, X.; Liu, K.; Long, Q.; Shen, L.; for the Alzheimer's Disease Neuroimaging Initiative. Deep multiview learning to identify imaging-driven subtypes in mild cognitive impairment. *BMC Bioinformatics* **2022**, *23*, 402. https://doi.org/10.1186/s12859-022-04946-x.

19. El-Sappagh, S.; Ali, F.; Abuhmed, T.; Singh, J.; Alonso, J.M. Automatic detection of Alzheimer's disease progression: An efficient information fusion approach with heterogeneous ensemble classifiers. *Neurocomputing* **2022**, *512*, 203–224. https://doi.org/10.1016/j.neucom.2022.09.009.

20. Emon, M.A.; Heinson, A.; Wu, P.; Domingo-Fernández, D.; Sood, M.; Vrooman, H.; Corvol, J.C.; Scordis, P.; Hofmann-Apitius, M.; Fröhlich, H. Clustering of Alzheimer's and Parkinson's disease based on genetic burden of shared molecular mechanisms. *Scientific Reports* **2020**, *10*, 19097. https://doi.org/10.1038/s41598-020-76200-4.

21. Wen, J.; Varol, E.; Sotiras, A.; Yang, Z.; Chand, G.B.; Erus, G.; Shou, H.; Abdulkadir, A.; Hwang, G.; Dwyer, D.B.; et al. Multi-scale semi-supervised clustering of brain images: Deriving disease subtypes. *Medical Image Analysis* **2022**, *75*, 102304. https://doi.org/10.1016/j.media.2021.102304.

22. Poulakis, K.; Pereira, J.B.; Muehlboeck, J.S.; Wahlund, L.O.; Smedby, O.; Volpe, G.; Masters, C.L.; Ames, D.; Niimi, Y.; Iwatsubo, T.; et al. Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. *Nature Communications* **2022**, *13*, 4566. Number: 1 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41467-022-32202-6.

23. Odusami, M.; Maskeliūnas, R.; Damaševičius, R. Optimized Convolutional Fusion for Multimodal Neuroimaging in Alzheimer's Disease Diagnosis: Enhancing Data Integration and Feature Extraction. *Journal of Personalized Medicine* **2023**, *13*, 1496. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, https://doi.org/10.3390/jpm13101496.

24. Dadu, A.; Satone, V.; Kaur, R.; Hashemi, S.H.; Leonard, H.; Iwaki, H.; Makarious, M.B.; Billingsley, K.J.; Bandres-Ciga, S.; Sargent, L.J.; et al. Identification and prediction of Parkinson's disease subtypes and progression using machine learning in two cohorts. *npj Parkinson's Disease* **2022**, *8*, 1–12. Number: 1 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41531-022-00439-z.

25. Su, C.; Hou, Y.; Xu, J.; Xu, J.; Brendel, M.; Maasch, J.R.M.A.; Bai, Z.; Zhang, H.; Zhu, Y.; Henchcliffe, C.; et al. Integrative analyses of multimodal clinical, neuroimaging, genetic, and transcriptomic data identify subtypes and potential treatments for heterogeneous Parkinson's disease progression, 2022. Pages: 2021.07.18.21260731, https://doi.org/10.1101/2021.07.18.21260731.

26. Nguyen, N.D.; Wang, D. Multiview learning for understanding functional multiomics. *PLOS Computational Biology* **2020**, *16*, e1007677. Publisher: Public Library of Science, https://doi.org/10.1371/journal.pcbi.1007677.

27. Huang, S.C.; Pareek, A.; Seyyedi, S.; Banerjee, I.; Lungren, M.P. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ Digital Medicine* **2020**, *3*, 136. https://doi.org/10.1038/s41746-020-00341-z.

28. Alayrac, J.B.; Donahue, J.; Luc, P.; Miech, A.; Barr, I.; Hasson, Y.; Lenc, K.; Mensch, A.; Millican, K.; Reynolds, M.; et al. Flamingo: a Visual Language Model for Few-Shot Learning, 2022. arXiv:2204.14198 [cs], https://doi.org/10.48550/arXiv.2204.14198.

29. Akbari, H.; Yuan, L.; Qian, R.; Chuang, W.H.; Chang, S.F.; Cui, Y.; Gong, B. VATT: Transformers for Multimodal Self-Supervised Learning from Raw Video, Audio and Text, 2021. arXiv:2104.11178 [cs, eess], https://doi.org/10.48550/arXiv.2104.11178.

30. Chen, R.J.; Lu, M.Y.; Weng, W.H.; Chen, T.Y.; Williamson, D.F.K.; Manz, T.; Shady, M.; Mahmood, F. Multimodal Co-Attention Transformer for Survival Prediction in Gigapixel Whole Slide Images. 2021, pp. 4015–4025.

31. Jack Jr., C.R.; Bernstein, M.A.; Fox, N.C.; Thompson, P.; Alexander, G.; Harvey, D.; Borowski, B.; Britson, P.J.; L. Whitwell, J.; Ward, C.; et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* **2008**, *27*, 685–691. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.21049, https://doi.org/10.1002/jmri.21049.

32. Machado-Reyes, D.; Kim, M.; Chao, H.; Shen, L.; Yan, P. Connectome transformer with anatomically inspired attention for Parkinson's diagnosis. In Proceedings of the Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, New York, NY, USA, aug 2022; BCB '22, pp. 1–4. https://doi.org/10.1145/3535508.3545544.

33. Reyes, D.M.; Kim, M.; Chaoh, H.; Hahn, J.; Shen, L.; Yan, P. Genomics transformer for diagnosing Parkinson's disease. In Proceedings of the 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), sep 2022, pp. 01–04. ISSN: 2641-3604, https://doi.org/10.1109/BHI56158.2022.9926815.

34. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer Normalization. *arXiv:1607.06450 [cs, stat]* **2016**. arXiv: 1607.06450.

35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems. Curran Associates, Inc., 2017, Vol. 30.

36. Fischl, B.; Dale, A.M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences* **2000**, *97*, 11050–11055.

37. Hartig, M.; Truran-Sacrey, D.; Raptentsetsang, S.; Simonson, A.; Mezher, A.; Schuff, N.; Weiner, M.; et al. UCSF freesurfer methods. *ADNI Alzheimers Disease Neuroimaging Initiative: San Francisco, CA, USA* **2014**.

38. Bellenguez, C.; Küçükali, F.; Jansen, I.E.; Kleineidam, L.; Moreno-Grau, S.; Amin, N.; Naj, A.C.; Campos-Martin, R.; Grenier-Boley, B.; Andrade, V.; et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics* **2022**, *54*, 412–436. Number: 4 Publisher: Nature Publishing Group, https://doi.org/10.1038/s41588-022-01024-z.

39. Buniello, A.; MacArthur, J. GWAS Catalog, 2019. Vol. 47 (Database issue): D1005-D1012.

40. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **2007**, *81*, 559–575. https://doi.org/10.1086/519795.

41. Pedregosa, F.; others. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.

42. Hand, D.J.; Till, R.J. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning* **2001**, *45*, 171–186. https://doi.org/10.1023/A:1010920819831.

43. Zhou, T.; Thung, K.H.; Zhu, X.; Shen, D. Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human Brain Mapping* **2019**, *40*, 1001–1016. https://doi.org/10.1002/hbm.24428.

44. Doody, R.S.; Massman, P.; Dunn, J.K. A Method for Estimating Progression Rates in Alzheimer Disease. *Archives of Neurology* **2001**, *58*, 449–454. https://doi.org/10.1001/archneur.58.3.449.

45. Doody, R.S.; Pavlik, V.; Massman, P.; Rountree, S.; Darby, E.; Chan, W. Predicting progression of Alzheimer's disease. *Alzheimer's Research & Therapy* **2010**, *2*, 2. https://doi.org/10.1186/alzrt25.

46. Prosser, A.; Evenden, D.; Holmes, R.; Kipps, C. Progression modelling of cognitive decline and associated FDG-PET imaging features in Alzheimer's disease. *Alzheimer's & Dementia* **2020**, *16*, e045900. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/alz.045900, https://doi.org/10.1002/alz.045900.

47. Gu, Z.; others. circlize implements and enhances circular visualization in R. *Bioinformatics* **2014**, *30*, 2811–2812.

48. Terada, S.; Sato, S.; Nagao, S.; Ikeda, C.; Shindo, A.; Hayashi, S.; Oshima, E.; Yokota, O.; Uchitomi, Y. Trail Making Test B and brain perfusion imaging in mild cognitive impairment and mild Alzheimer's disease. *Psychiatry Research: Neuroimaging* **2013**, *213*, 249–255. https://doi.org/10.1016/j.pscychresns.2013.03.006.

49. Matías-Guiu, J.A.; Cabrera-Martín, M.N.; Valles-Salgado, M.; Pérez-Pérez, A.; Rognoni, T.; Moreno-Ramos, T.; Carreras, J.L.; Matías-Guiu, J. Neural Basis of Cognitive Assessment in Alzheimer Disease, Amnestic Mild Cognitive Impairment, and Subjective Memory Complaints. *The American Journal of Geriatric Psychiatry* **2017**, *25*, 730–740. https://doi.org/10.1016/j.jagp.2017.02.002.

50. Camacho, J.; Rábano, A.; Marazuela, P.; Bonaterra-Pastra, A.; Serna, G.; Moliné, T.; Ramón y Cajal, S.; Martínez-Sáez, E.; Hernández-Guillamon, M. Association of CD2AP neuronal deposits with Braak neurofibrillary stage in Alzheimer's disease. *Brain Pathology* **2021**, *32*, e13016. https://doi.org/10.1111/bpa.13016.