Alzheimer's subtypes
A supervised, unsupervised, multimodal, multilayered embedded recursive (SUMMER) AI study

S. Kinreich[1], A. Bingly[1], G. Pandey[2]

[1]Department of Psychiatry and Behavioral Sciences, SUNY Downstate Health Sciences University, Brooklyn, NY, USA.
[2]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract

Since Alzheimer's disease (AD) is a heterogeneous disease, different subtypes may have distinct biological, genetic, and clinical characteristics, requiring tailored interventions. While several proposed subtypes of AD exist, there is still no clear consensus on a definitive classification. By leveraging complementary AI approaches, including supervised and unsupervised learning, within a recursive pipeline (SUMMER) that integrates multimodal datasets encompassing MRI measurements, phenotypes, and genetic data, our goal was to generate robust scientific evidence for identifying AD subtypes. Data was downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and included neuroimaging data (MRI), genetics (SNPs), clinical diagnosis, and demographics. 1133 European American participants' images, aged 55-95, were included in this study. The analysis was multi-fold, where the first step involved applying an unsupervised application to a subset of the MRI sample (AD + cognitively normal (CN) aged matched groups, 100 men aged 68–85 years, and 76 women aged 68–85 years). The MRI brain gray matter was segmented into 44 regions of interest (ROIs) according to a standard atlas, and 618 features were extracted, including ROI voxel intensity measurements such as minimum, maximum, and histogram variables. Results identified a cluster of subtype AD men and a cluster of subtype AD women that were distinct from the rest of their respective samples. In the next step, the integrity of the identified subtype AD clusters was investigated using the XGBoost supervised machine learning application with genetic features (SNPs, N=36,724) and labels: the identified subtype AD cluster vs. the rest of the sample, stratified by sex. A significant AD subtype men model (accuracy=0.85, F1=0.72, AUC=0.83) and a significant women AD subtype model (accuracy=0.81, F1=0.81, AUC=0.81)

were built, confirming the homogeneity of the isolated AD subtype clusters. Discriminative biomarkers were extracted from the significant models, including selected ROIs and SNPs. Finally, the subtype models were tested on an unseen subset of ADNI data. The genetic-based models identified clusters of AD subtype participants consisting of 34% of the men AD group and 47% of the women AD group. Phenotypic analysis indicates that lower body weight was associated with the women's AD subtype. Complex diseases like AD demand a sophisticated, multimodal approach for precise diagnosis. Effectively identifying disease subtypes enhances the potential for personalized treatment, ultimately improving patient outcomes.

Introduction

Cognitive decline in older adults is not attributed to a single neurological process but rather to several distinct subtypes[1]. One subtype is associated with natural aging[2], while others are associated with neurological conditions such as Alzheimer's disease (AD) and related dementia[3]**.** These diverse subtypes create ambiguity in diagnosing AD and understanding its origins, posing significant challenges for developing effective treatments and therapies[4]. It was estimated that in 2022, around 6.5 million Americans aged 65 and older were living with AD[5]. US forecasts indicate that the population of Americans with either Alzheimer's dementia or mild cognitive impairment (MCI) will reach 15 million by 2060, up from 6.08 million in 2017[6]. These estimates highlight the critical importance of identifying risk and resilience biomarkers associated with each subtype's underlying origins, enabling the personalization of future treatments and prevention strategies.

To this end, our work focused on phenotypic and biomarker differences between cognitive normal (CN) individuals and AD patients. We aim to build classifiers and predictive models to identify AD subtypes, distinguishing between AD and CN groups and subgroups within the groups to identify biomarkers associated with each.

Previous studies investigating AD subtypes suggested that based on pathologic measures of senile plaques, neurofibrillary tangles (NFTs, an abnormal accumulation of a protein called tau inside neurons)[7,8], and cerebral amyloid angiopathy (CAA, a protein build-up on the walls of the arteries in the brain)[9], 3 subtypes can be defined; typical AD, with balanced NFT counts in the hippocampus and association cortex[8]; limbic-predominant AD, with counts predominantly in the hippocampus[8]; and hippocampal-sparing AD, with counts predominantly in the association cortex[7,8]. Several neuroimaging studies have presented findings that support these subtypes by using visual rating scales of brain atrophy or automated methods for estimating regional brain

volume[10,11]. A fourth subtype displaying minimal brain atrophy was also identified by neuroimaging studies[8]. Genetic studies have also explored the possibility of defining AD through distinct subtypes. For example, the MAPT H1H1 (microtubule-associated protein tau (MAPT) H1H1 genotype) was suggested to have a higher frequency in the limbic-predominant AD subtype compared with typical and hippocampal-sparing AD.[10,12] However, another study reported that typical AD had the highest frequency of MAPT H1H1 carriers, followed by limbic-predominant and hippocampal-sparing AD[13]. The APOE gene E4 allele (APOE4) has been extensively studied and recognized as a major risk factor for late-onset AD[14]. Additionally, individuals who inherit two copies of the APOE4 variant face a substantially heightened risk of developing the condition[14]. With these encouraging results, the search for distinct AD subtypes continues, focusing on developing reliable and accurate diagnostic tools based on brain structure and function, genetics, and other phenotypes to identify AD subtypes effectively.

The complexity of identifying AD poses challenges when different subtypes are associated with different biological markers[15]. At the same time, biomarker analysis based solely on diagnosis and symptoms may be biased because symptoms are often similar between subtypes[15]. Supervised machine learning algorithms may produce biased results due to misleading labeling that fails to account for the diversity of AD subtypes. Furthermore, the process of healthy aging can vary significantly in terms of brain anatomy and the timeline of changes, and this variation may also have an effect when interacting with disease subtypes[16]. To that end, unsupervised clustering can be effective, especially in the early stages when symptoms have not fully developed[17]. In the unsupervised approach, the software does not rely on prior knowledge that differentiates between groups, but rather looks for consistent patterns in the data to create clusters[18]. Clustering is a data-driven technique of grouping based on features that are similar to each other and separating them from those that are different[19]. Clusters are formed with features having minimum intra-cluster distance and maximum inter-cluster seperation[19]. An additional strength of the unsupervised learning approach is that it suffers less from data overfitting due to its embedded nature, where learning is done in an entirely blind manner[18]. Combining unsupervised and supervised methodologies may leverage the strengths of both approaches to effectively identify clusters within a dataset.
Moreover, our groundbreaking research[20] and others[21] have demonstrated that multimodal AI algorithms often outperform single-modality models in classifying complex diseases. By integrating diverse domains, such as environmental factors, genetics, and psychiatric diagnoses, our work leverages the strengths of each field while addressing their individual

limitations, offering a more comprehensive framework for disease understanding and classification[20]. AD is a complex condition with potentially diverse origins, necessitating a multifaceted approach to identify and understand its various subtypes. In this study, we aimed to integrate multiple complementary approaches and datasets, including supervised, unsupervised, multimodal, multilayered embedded recursive (SUMMER) machine learning algorithms, brain structure, genetics, and phenotypic analysis, to uncover subtypes of AD.

To address this study's aims, we used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI)[22,23]. 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 MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD[23]. ADNI is a multisite, longitudinal study employing imaging, clinical, bio-specimen, and genetic biomarkers in CN as well as in individuals with early MCI and those who are diagnosed with AD[24]. ADNI follows individuals with and without AD and MCI, enabling a unique opportunity to assess an individual's health status while incorporating age, sex, and ancestry differences into the analysis[22].

To summarize, with ADNI's rich genetic and phenotypic data from multiple domains (e.g., clinical, neuroimaging (MRI), psychosocial, psychiatric, and demographic), we applied brain-based unsupervised algorithms that have the potential to identify biodiversity patterns related to dementia subgroups that would otherwise remain hidden. We utilized the ADNI's individual genetics information to further confirm the AD subtype's robustness initially found by the clustering of the unsupervised machine learning algorithm. Our multilayered analysis continued and validated the results with recursive analysis, where the AD subtype finding was validated with unseen ADNI data.

Method

We present here an assembled pipeline constructed from compatible machine learning applications (unsupervised and supervised) that were applied to various modalities (neuroimaging, genetics, and phenotypes) in order to identify and establish distinctive AD subtypes. All calculations were done using Python packages and FreeSurfer (version 4, http://surfer.nmr.mgh.harvard.edu, Boston, MA)[25].

ADNI Sample

The MRI data acquired from AD and CN participants were downloaded from the ADNI database (https://adni.loni.usc.edu/)[26]. Only European American (EA) participants were included in the analysis due to the low number of other ancestries in the ADNI dataset, and only participants with genetic information, resulting in an initial dataset that included 1,113 brain images of participants between the ages of 55.1 and 95.8 years. Initial data included multiple visits by each participant (588 men's brain images and 525 women's brain images). Details of the ADNI design, participant recruitment, clinical testing, and additional methods have been published previously[22,27] and are available at www.adni-info.org.

MRI scans

Raw ADNI 1.5 T and 3.0 T MRI scans were downloaded from the ADNI public database (www.loni.ucla.edu/ADNI). The raw DICOM files were transformed into NIFTI format[25].

MRI preprocessing region segmentation

The MRI data was preprocessed using standard procedures that included 1**.** skull-stripping using the Brain Extraction Tool (BET)[28] (Fig 1A). 2. Brain normalization to standardize all brains to the same scale. 3. Brain registration to the standard template (MNI152_T1_1mm_brain) using the Python package of ANTS[29]. This process harmonized the data into a voxel resolution of $1mm^3$. 4. Brain segmentation according to the structural standard brain template, defining 44 ROIs for each brain[30]. 5. The ROIs masks were created using The FSLView[31] with the same brain template (MNI152_T1_1mm_brain) and atlas (Harvard-Oxford cortical and subcortical structural atlases)[30] (Fig 1B). ROIs included cortical and subcortical structural areas according to the atlas[30] (full list in the supplementary file). Only selected available ROIs were extracted separately from the left and right hemispheres (amygdala, thalamus, putamen, hippocampus, caudate, and nucleus accumbens), while the rest were averaged across the hemispheres.

MRI ROI measurements

The most straightforward way to examine brain anatomy is through visual inspection of MRI scans[32,33]. However, this method is time-consuming, requires substantial operator training, and suffers from inconsistent tracing protocols across laboratories[32]. Moreover, research suggests that visual inspection may be subject to perceptual bias[34]. To address these challenges, automated software tools that analyze MRI intensities have been developed to assist in identifying tumors[35], delineating ROI volumes[36], and quantifying other brain variables[37]. MRI

voxel intensities are rarely used directly as features in neuroimaging research, with only a small subset of Alzheimer's disease studies employing voxel-based intensity analyses[38-40]. Compared to functional MRI (fMRI), voxel intensity analysis does not require complex task-based experiments and can be applied to any standard structural MRI scan[31]. Following the preprocessing and harmonization/normalization steps that allowed the MRI scans to be on the same scale, the following measurements of the MRI voxel values were calculated: minimum, maximum, and standard deviation. Additional measurements included values within the bottom 10th, 20th, and 25th percentiles, as well as the top 10th, 20th, and 25th percentiles. These values were calculated to reduce the possibility that ROI's intensity outliers will bias the results[41]. Additionally, histograms were calculated for each ROI's (Fig 1C) voxel values, and the histogram entropy was added to the features (see Table 1 in supplementary file for feature list). Next, for selected ROIs (amygdala, thalamus, putamen, hippocampus, caudate, and nucleus accumbens), the difference between the contralateral histogram ROIs was calculated. Functional and morphological differences between hemispheres have previously been suggested in the study of various brain disorders[42]. Contralateral measurements included the

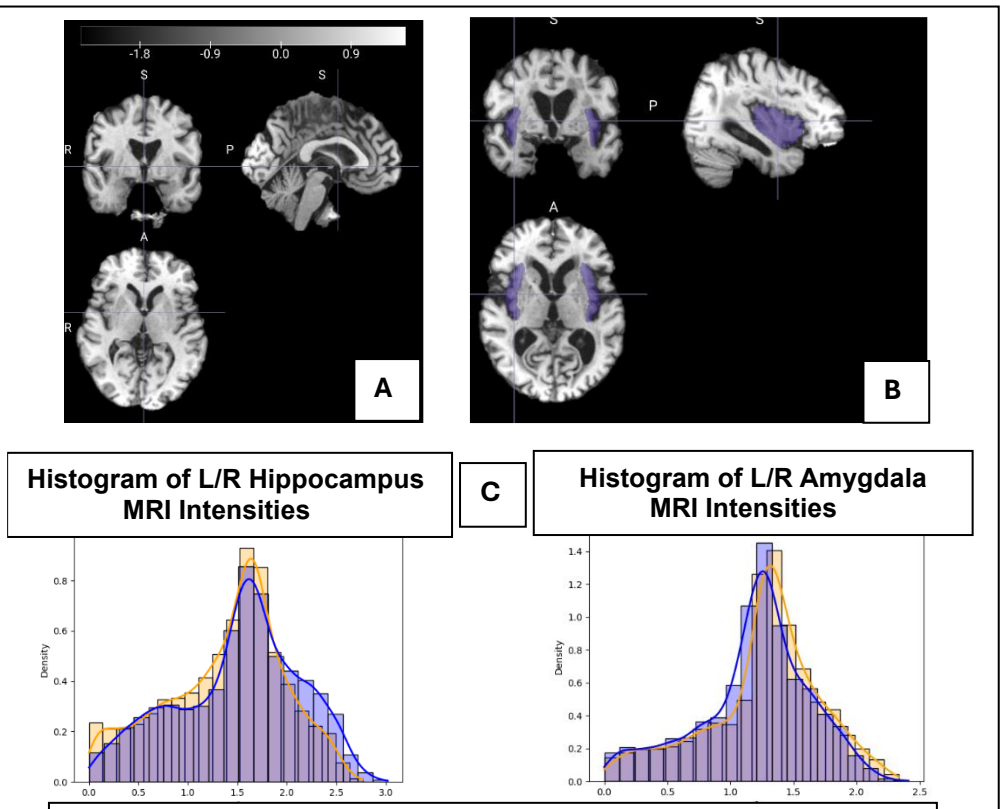

Figure1. Brain preprocessing and ROI extraction. A. The brain image was stripped from the scalp, registered to a standart template, and normalized. B. 44 ROIs intensities values were extracted from each brain. Here is an illustration of the extracted insula ROI colored in purple. C. Histogram of intensities of the contralateral Hippocampus (on the left) and Amygdala (on the right). Right ROI, Orange. Left ROI, blue and shared area; purple.

Euclidean distance, which is the difference between the entropy of histograms, the percentage of mutual area (vs. both histograms), the percentage of mutual areas (vs. the right), and the percentage of mutual areas (vs. the left) (Fig 1C). Overall, the analysis included 618 features for each ROI. The features were randomly divided into 36 groups of nine to maximize their individual contributions[43] and the ML algorithm iterated over these groups. A cluster was defined as visibly isolated from the other participants and included more than 85 percent of AD participants. This threshold was chosen arbitrarily to ensure the inclusion of the majority of the AD participants in the cluster.

Genetics

Raw genetic information was downloaded from the ADNI public database (www.loni.ucla.edu/ADNI). The Human 610-Quad BeadChip (Illumina, Inc., San Diego, CA) was used to genotype all ADNI participants in the database, which resulted in a set of 620,901 SNPs[44] for each. The specific methods related to ADNI genotyping can be found on the ADNI website (https://adni.loni.usc.edu/data-samples/adni-data/genetics-related-omics).
To further study the homogeneity of the MRI-based isolated cluster as an AD subgroup, we investigated the genetic profile of the cluster compared to the rest of the group. Supervised machine learning analysis (XGBoost) was applied with SNPs as features and with the isolated group and the rest of the group as labels. The SNPs were first normalized, and then SNPs with missing values (more than 5% for men, 10% for women) were discarded. Next, we used the least absolute shrinkage and selection (LASSO) method for feature selection[45]. The sparsity property of LASSO (i.e., most coefficient estimates being exactly zero) makes it attractive for feature selection, as it reduces the estimated variance while providing a more interpretable final model[46]. Its applications to genomic data[47] and behavioural markers[48] have shown that selecting a small number of representative features can achieve more accurate classification than all the features. We first determined the LASSO regularization parameter using a 10-fold cross-validation (CV) procedure with labels "isolated cluster" and "rest of the group" as the response variable. All the features with a non-zero coefficient were retained for subsequent analyses.

Demographic and clinical data

Demographic and diagnosis information for all analyzed visits was downloaded from the ADNI clinical data repository (https://www.loni.ucla.edu/ADNI/Data/ADCS_Download.jsp).

SUMMER Analysis Pipeline

The outline of the AD subtype pipeline (including supervised AI, unsupervised AI, multimodal, multilayered embedded recursive approaches (SUMMER) is presented in Figure 2.

Following the SUMMER pipeline, we conducted a series of AI analyses, each incorporating different samples and modalities, with each analysis building upon the results of the previous one. We first identified the AD clusters using an exploratory unsupervised approach with MRI data. To ensure the integrity of the clusters, we then validated them using a supervised approach incorporating genetic data. Finally, we tested the models on an unseen ADNI data subset to ensure robustness and generalizability. Below is the breakdown of the steps.

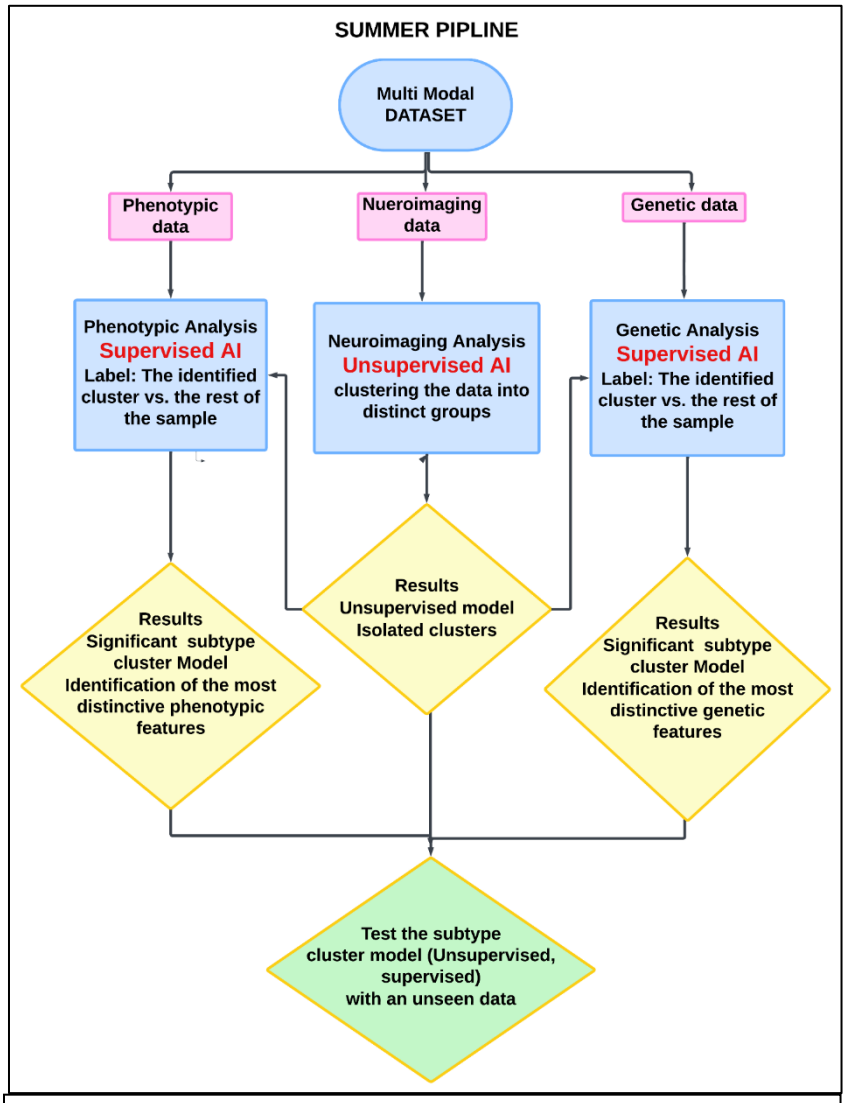

Figure 2. SUMMER pipline

1. **Unsupervised machine learning clustering with structural (MRI) features**

AI unsupervised learning applications in machine learning allow the interpretation of data without predefined labels[19]. The applications rely on algorithms that identify patterns, structures, and relationships in the data[18]. Our sample included ADNI Age-matched AD and CN participants, stratified by sex (men, women), ranging from 68 to 85 years old[49] (See Table 1 for participant information). The groups (AD, CN) included two visits from each participant to increase the sample size. The age range constraint was applied based on previous findings, confirming age's significant impact on brain structure and function[50]. Similarly, previous findings

Table 1. The Unsupervised Analysis Sample					
	# CN images	# AD images	Mean Age CN ± SD	Mean Age AD ± SD	P value
Men	50	50	78.12±4.37	77.83±4.85	0.74
Women	38	38	76.53±4.18	76.52±4.19	0.99
CN: Cognitively normal; AD: Alzheimer's Disease; SD: standard deviation					

on sex differences in brain function were the reason for stratified analysis by sex[51]. A total of 618 features were computed per ROI (see Methods). Given 44 ROIs per participant, this resulted in 27,192 features per participant. The dimensionality of the features was reduced by performing a t-test on all features, comparing AD to CN to focus on AD-related characteristics. Only features that met the significance threshold ($p < 0.05$) were included in the unsupervised analysis, resulting in a final set of 50 MRI features. We applied the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP)[52] unsupervised algorithm with the MRI features. UMAP[53] is a software tool for unsupervised dimensionality reduction, which is pivotal for visualizing and clustering unannotated high-dimensional data[53]. Dimensionality reduction is aimed at transforming the data into a relevant and reduced dimensional space by discovering intrinsic data structure and limiting redundancy. It allows extracting useful variables while discarding noise and correlated features[53]. UMAP is a manifold learning technique based on Riemannian geometry and algebraic topology. UMAP was used with the participant's MRI ROI values to capture the underlying patterns. The ROIs' features that contributed most significantly to cluster formation were identified by performing a t-test between the cluster and the rest of the data for each significant model (men and women). A false discovery rate (FDR) was applied to correct the results for multiple comparisons[54].

2. Supervised machine learning with genetic data

Here we aimed to reinforce the AD cluster and its integrity further by investigating its genetic homogeneity compared to the rest of the sample, which includes both CN and AD participants. The sample included the same cohort from step 1, i.e., age-matched AD and CN participants, stratified by sex, ranging from 68 to 85 years old, was included in this analysis.
Features included ADNI's genetic data (SNPs)[55] Following the LASSO application, all the features with non-zero coefficients were retained for subsequent analyses, resulting in ~20 SNPs for men and ~20 SNPs for women.

The Supervised eXtreme Gradient Boosting (XGBoost)[56] was applied with the isolated cluster vs. the rest of the group as labels. XGBoost is an open-source machine learning library that uses gradient-boosted decision trees[56]. It is known for being fast, efficient, and able to handle large and small datasets[57]. The reduced set of features identified was fed to the XGBoost algorithm to classify the isolated cluster.

3. Testing the genetic-based model with ADNI unseen data

In this step, the cluster model built by the genetic-based supervised algorithm was tested on an independent dataset comprising all ADNI participants who were not included in the previous analysis and did not serve as the model's foundation (N men =365, N women=438) with age range of 55.1 and 95.8. The sex-specific models were applied with their corresponding sample and features (SNPs)(more information about the sample is given in the supplementary file). The model was expected to identify a subtype of AD that would include AD patients.

4. Phenotype analysis

The subtype clusters were also tested for phenotypic (body weight) association with AD. Evidence suggests that body weight can impact the risk and progression of AD. Being overweight or obese in midlife is linked to a higher risk of developing AD later[58]. Excess weight can contribute to conditions like high blood pressure, diabetes, and inflammation, all of which are associated with cognitive decline[59]. On the other hand, People with AD often experience unintentional weight loss, especially in the later stages[60]. This can be due to decreased appetite, difficulty eating, changes in metabolism, or trouble recognizing food[60]. To test the relationship between AD and body weight, a t-test was conducted between the recorded weight data of the participants in the cluster and the weight data of the rest of the sample.

Results

1. AI Unsupervised algorithm

We applied UMAP[52] unsupervised algorithm with the MRI features. All results (distribution of the participants' UMAP values in 2d space) from the unsupervised analysis can be found in the supplementary file. Results for the women group revealed an isolated cluster composed mainly of AD participants (86% AD, Figure 3A). Results for the men group revealed an isolated cluster composed mainly of AD participants (92.3% AD, Figure 4A). The features that most strongly contributed to the formation of the distinct cluster of men and women AD subtypes were found

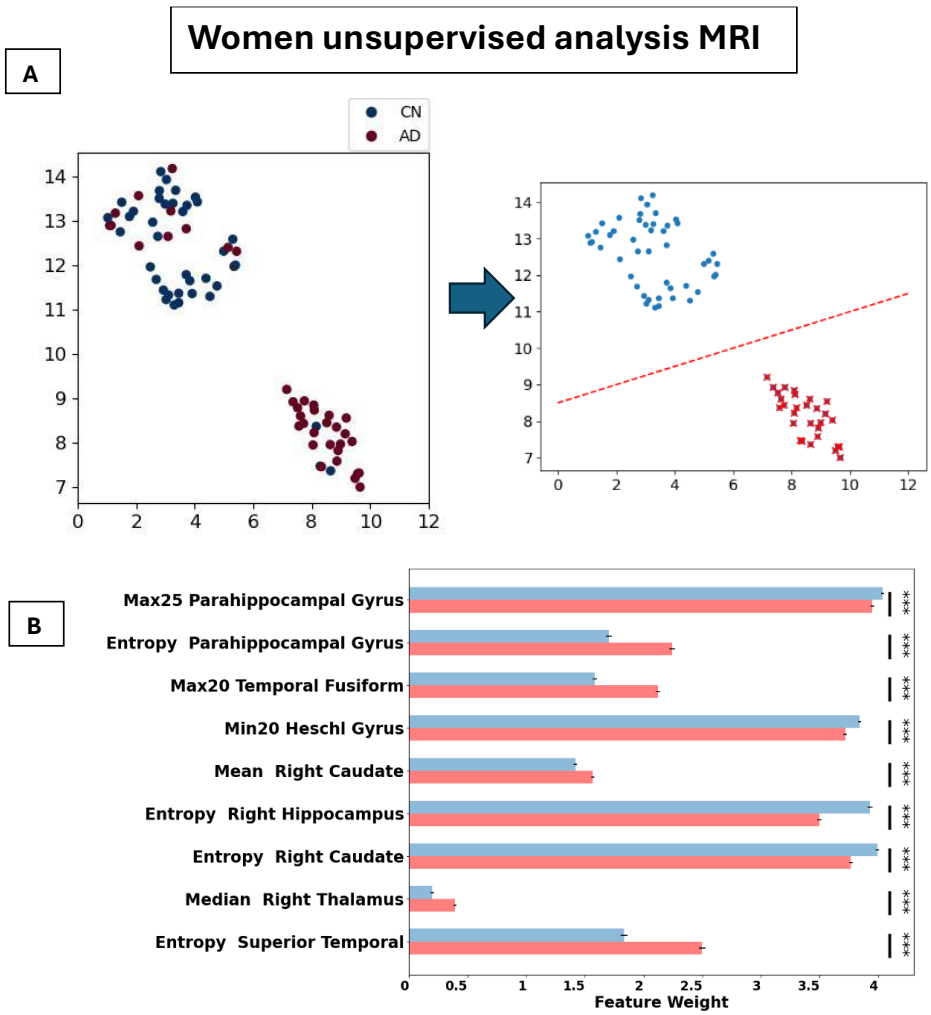

Figure 3. The women's unsupervised clustering of the MRI measurements revealed an isolated AD cluster. A. It is evident that the majority of the isolated cluster includes AD participants. On the right, the isolated cluster is colored in red, and the rest of the sample is colored in blue. B. The bar graph presents ROIs that contributed the most to the creation of the cluster.
Max25-top 25th percentile, Max20-top 20th percentile, Min20-bottom 20th percentile,

to be primarily associated with brain networks known to underlie AD-related memory function, including regions such as the hippocampus and parahippocampus. The fusiform gyrus, traditionally associated with face recognition processes, was also found to be linked to both sexes AD subtypes. The women AD subtype was found to be associated with the thalamus and the caudate and men-specific AD subtype was associated with the parietal operculum (Figure 3B (women) and 4B (men)). Overall, all MRI average values were lower for the AD cluster subtype and higher for the SD values and entropy values for men and women. The only ROIs with higher AD subtype MRI values were the Supracalcarine Cortex (men), Parietal Operculum (men) and Lingual Gyrus (men). Table 2 in the supplementary file presents the isolated cluster

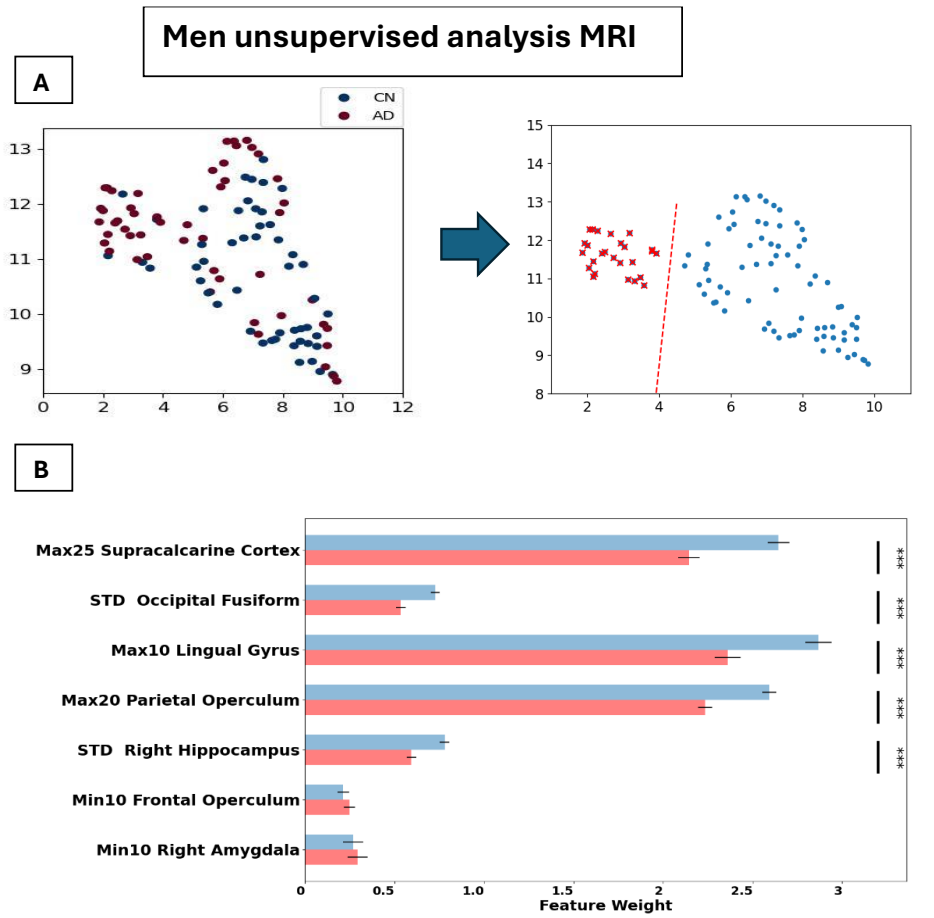

Figure 4. The men's unsupervised clustering of the MRI measurements revealed an isolated AD cluster. A. It is evident that the majority of the isolated cluster includes AD participants. On the right, the isolated cluster is colored in red, and the rest of the sample is colored in blue. B. The bar graph presents the ROIs that contributed the most to the creation of the cluster. Max25-top 25th percentile, Max20-top 20th percentile, Max10-top 10th percentile, Min10-bottom 10th percentile. STD – Standard deviation

vs. the rest of the data mean values and p-values, and Table 3 in the supplementary file presents the mean values and the p-values of all brain areas associated with the clusters.

2. Supervised machine learning with genetic data

Significant models were created for both men (accuracy=0.85) and women (accuracy=0.81). Table 2 presents model information, including precision, recall, F1 score, and area under the curve (AUC), confirming the genetic heterogeneity of the clusters. The most important features contributing to the formation of the clusters include SNPs such as; (men) rs7722673 (gene

WWC1), rs1538608, rs2396274, rs1122269, and rs17131515 (gene CDH4), and (women) rs11689050 (gene LINC01808) rs1874161, rs10216685, rs6775796, rs2422973 (gene RASSF2), rs12130821 (gene RGS5), rs9425322 (gene RGL1), rs17794576 (gene NCAM2), rs7306261 (Figure 5). A full, detailed SNPs summary is presented in Supplementary Table 4.

Table 2. XGBoost Supervised AI Model Cluster Accuracy

Sex	Precision	Recall	Accuracy	F1	AUC
Men	0.66	0.8	0.85	0.72	0.83
Women	0.75	0.85	0.812	0.81	0.81

3. Testing the genetic-based model with ADNI unseen data

The genetic-based models (men and women) were tested with unseen data. Figures 6A and 6B show the sample percentage of AD subtype clusters that were identified by the men and women models, respectively. Results show that 47% of the women subtype cluster comprised AD

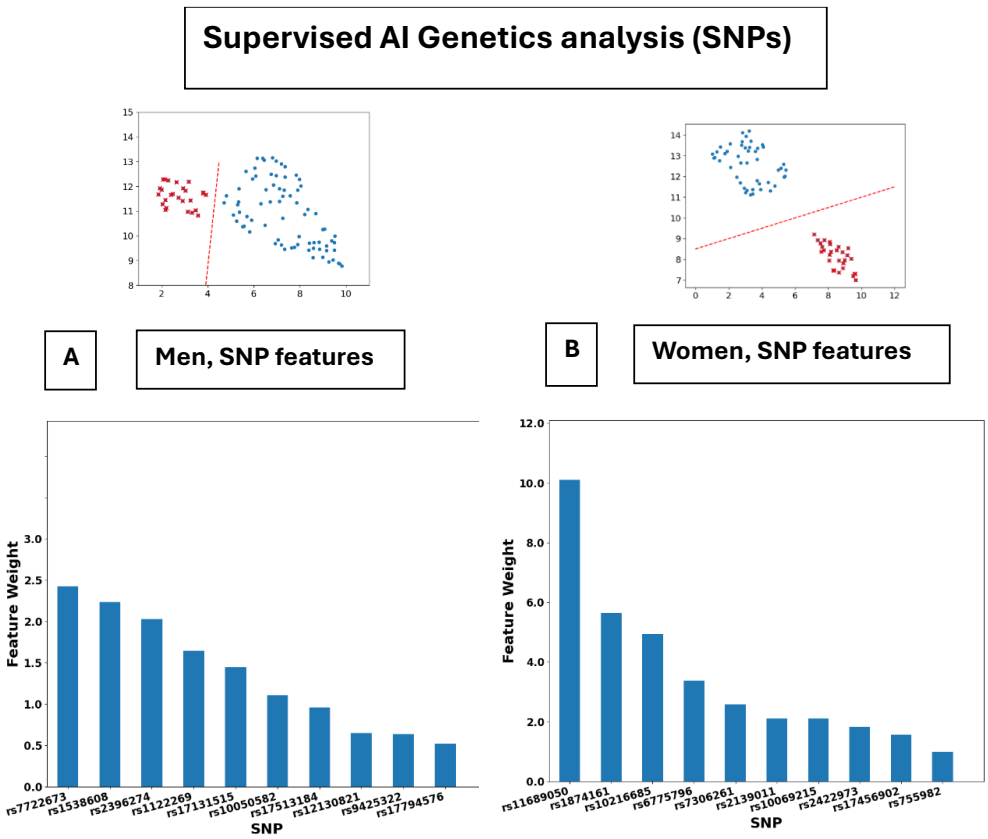

Figure 5. XGBoost Supervised AI Genetics Analysis (SNPs). The XGBoost analysis labels were the cluster (colored in red) vs. the rest of the sample (colored in blue). A. SNPs that are the most significant contribution to the XGBoost model for men (A) and women in (B)

participants, while in the rest of the data included 10% AD participants (Figure 6). 30% of the men cluster were AD, while 15% were AD in the rest of the sample (Figure 6 and see Table 5 in supplementary file).

4. Phenotype analysis

Results show an association between the participant's weight and the women AD subtype cluster, p = 0.004 (AD, mean=63.29±9.85, CN, mean=71.31±13.65). No association was found between men AD subtype cluster and body weight, p == 0.6 (AD, mean = 83.54±9.82, CN, mean=82.14±13.13 (Table 3. for full subtype biomarkers and phenotypic description)

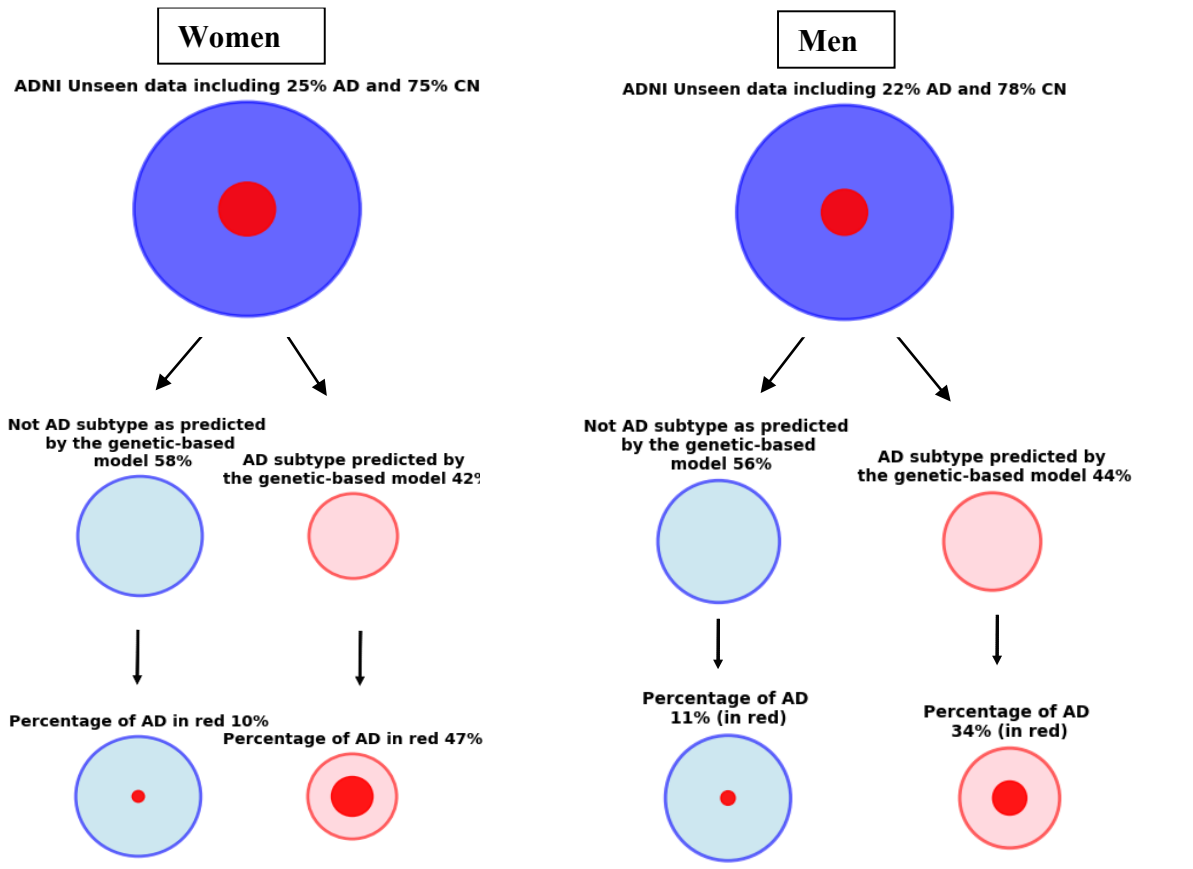

Figure 6. Genetics-based AD subtype model applied with ADNI unseen dataset. For both sexes, the models identified a cluster that included AD patients.

Table 3: AD subtype of women and men, selected features		
	Women AD subtype	**Men AD subtype**
ROI	top 25th Parahippocampal Gyrus	top 25th Supracalcarine Cortex
	Entropy Parahippocampal Gyrus	STD Occipital Fusiform
	top 20th Temporal Fusiform	top 10th Lingual Gyrus
	bottom 20th Heschl Gyrus	top 20th Parietal Operculum
	Mean Right Caudate	STD Right Hippocampus
	Entropy right Hippocampus	
	Entropy Superior Temporal	
SNPs	rs11689050	rs7722673
	rs1874161	rs1538608
	rs10216685	rs2396274
	rs6775796	rs1122269
	rs7306261	rs17131515
	rs2139011	rs10050582
	rs10069215	rs17513184
	rs2422973	rs12130821
	rs17456902	rs9425322
Body weight	**Lower body weight**	

Discussion

Diagnosing AD subtypes, especially in the early stages of the disease, holds enormous potential for research on drug interventions and treatments. By mining high-content data sets using an exploratory, unsupervised data analytics approach, we were able to distinguish AD subgroups and identify their respective biomarkers. The world of medicine strives for more personalized medicine, identifying the needs of the individual over the "average" treatment in order to improve treatment results. In our innovative SUMMER model analysis, we have assembled different types of AI and multimodality approaches, incorporating multilayered and recursive steps enabled by the multimodal ADNI dataset. We identified distinctive subtype clusters of men AD and women AD by an unsupervised analysis applied to MRI intensity measurements. The most discriminative features in the AD subtype models include MRI measurements from ROIs, such as the hippocampus and the fusiform gyrus. The AD subtype cluster homogeneity was

confirmed with a significant genetics-based (SNPs) XGBoost supervised model. AD subtype model-specific SNPs were identified for each of the sexes. Phenotypic analysis of body weight found that the women AD subtype has lower body weight than the rest of the sample. This association was not found for the men's AD subtype. The significant genetic models were tested with ADNI unseen data identifying a men AD subtype cluster and women AD subtype cluster in this dataset. The results highlight the superiority of multimodality biomarkers in identifying subtypes of complex diseases such as AD.

Multi-modality features and AD subtypes

In this study, we demonstrated that utilizing multiple AI approaches and modalities allowed the identification of homogeneous clusters. Specifically, the unsupervised approach successfully classified AD subtypes that would have otherwise remained hidden within the averaged results. Our previous groundbreaking study on predicting the development of Alcohol Use Disorder[20,61], along with others[21], demonstrated that models integrating multimodal data, such as neuroimaging biomarkers and genetic information, outperform those relying on a single modality[20]. These findings align with the understanding that complex diseases are influenced by multiple factors, including genetic and environmental components[62]. Our results indicate that distinct AD subtypes can be effectively identified using a series of AI algorithms applied sequentially, and in a recursive manner across multiple datasets. An alternative AI approach explored in recent studies, also utilizing multi-modal data, is implemented as a single-layer model rather than a sequential one. In the one-layer approach, multiple features from different modalities are reduced to display an efficient set of features to predict/classify the labels. This reduction due to collinearity may cause a loss of information. For example, when building a model that is composed of multiple modalities, such as phenotypes, brain functions, and genetics, the trained model may be based on only some of the modalities. Additionally, certain modalities are naturally more sensitive to factors such as age and sex. By adopting a sequential approach, we can more effectively balance and manage the trade-offs imposed by these constraints. For example, in our study, we applied unsupervised algorithms to neuroimaging data across a restricted age range, which was necessary due to the significant impact of age on brain structure.

AD subtype-related brain features

AI unsupervised analysis revealed a set of ROIs associated with AD subtypes in both sexes. The identification of subtype-specific features, including brain ROI measures and genetic

factors, can provide valuable insights into the onset and progression of AD. Our findings show that men and women AD subtypes have been found to be associated with known AD-memory-related brain networks, such as the hippocampus and parahippocampal gyrus, located in the medial temporal lobe[63]. These brain structures play a crucial role in forming and consolidating new explicit (declarative) memories (e.g., facts and events)[64]. The hippocampus is consistently implicated in AD-related brain structure, function, and behavioral studies, showing reduced volume[65], rapid tissue loss[66] and functional disconnection with other brain regions[67]. The fusiform gyrus, traditionally associated with face recognition processes, was also found to be linked to AD subtypes in both sexes. The fusiform gyrus is located below the lingual and parahippocampal gyri, both also associated with the AD subtypes, and together, they have been highlighted in imaging studies, demonstrating a strong link to memory impairment[68,69]. Fusiform may shrink in people who develop AD[70] and genes within the fusiform gyrus may be involved in AD onset and progression[71]. The fusiform gyrus was also implicated in mild cognitive impairment and loss of facial recognition[72], both of which are associated with the development of AD. The women AD subtype was found to be associated with the thalamus and the caudate. The thalamus functions as a relay center, transmitting sensory and motor signals to the cerebral cortex and the caudate nucleus[73], a crucial part of the basal ganglia, involved in processing spatial information and regulating motor behavior[73]. Both brain structures were previously implicated in AD, showing volume reduction[74]. Furthermore, a recent imaging study revealed that the amyloid imaging marker AV-45, a PET imaging agent used to visualize beta-amyloid plaques in the brain, is elevated in the caudate nucleus and putamen of individuals experiencing rapid cognitive decline[75]. More women-specific AD subtype brain regions include the Heschl's gyrus. Heschel's gyrus processes the frequency, duration, and intensity of sounds and changes in hearing, especially speech perception in noise, all of which may be a preclinical marker of neurodegeneration.[76] More so, this brain structure was found to be associated with AD in a volumetric MRI study[77]. Men-specific AD subtype was associated with the parietal operculum, a parietal lobe region. The parietal lobe was implicated in AD, and its dysfunction contributes to memory, language, and visuospatial deficits, particularly in the early stages[78]. In a transcranial magnetic stimulation(TMS) study, the parietal operculum specifically has been found to be involved in proficient memory-guided haptic grasping[79]. This finding and others[80] led to the suggestion that the parietal operculum acts as a “hub,” where several sensory-motor streams originating from different cerebral areas converge[80]. Our results add to these accumulating AD-related findings, suggesting that this group of ROIs characterizes the specific AD subtype revealed in the unsupervised analysis.

AD subtype-related genetics

The results show that the AD subtype clusters were genetically identified by a collection of SNPs, confirming the subtype's generalizability. Women-specific SNPs included a locus on chromosome 2, gene Long Intergenic Non-Protein Coding RNA 1808 (LINC01808) (rs11689050, weight ranking 1). In GWAS analyses, LINC01808 has been found to be associated with phenotypes such as height[81], pulse pressure measurement[82], ascending thoracic aortic diameter[83], and directly relevant to the current study, AD (MTAG)[84], and PHF-tau measurement[85]. PHF-tau, or paired helical filament tau, is a biomarker for AD, and its measurement involves detecting hyperphosphorylated tau protein, which forms neurofibrillary tangles[86]. The AD women subtype was also associated with a locus on chromosome 20, gene RASSF2 (rs2422973, weight ranking 8). This gene is associated with DNA methylation variation (age effect)[87] and encodes a protein that contains a Ras association domain (rheumatoid arthritis (RA) protein)[88], which plays a crucial role in signal transduction pathways (EGLN1/RASSF2 protein level ratio)[89]. The men AD subtype was associated with a locus on chromosome 5, gene WWC1 (rs7722673, weight ranking 1). A GWAS study found that WWC1 gene is associated with protein KIBRA levels[90]. KIBRA is a protein involved in learning and memory, and its levels are associated with cognitive impairment and tau pathology in conditions like AD[90]. WWC1 was also found to be associated with brain morphology in a schizophrenia study[91] and educational attainment[92]. Another locus was found on chromosome 1, gene CDH4 (rs17131515, weight ranking 5). CDH4 encodes a protein called retinal cadherin (R-cadherin) which is a cell adhesion protein that plays a crucial role in maintaining the structural integrity of tissues and organs, particularly in the retina and brain[93]. Studies show an association of this gene with educational attainment[94]. Both men and women subtype clusters found genetic association with smoking initiation (CDH4 and LINC01808 genes, respectively).[95] A locus on chromosome 1, RGS5 gene (rs12130821, weight ranking 8) has been associated with height in a diverse ancestry GWAS[81]. In the same GWAS, gene RGL1 (rs9425322, weight ranking 9) was found to be associated with height. In a GWAS of both sexes, RGS5 was associated with testosterone in the men group[96]. In a VA Million Veteran Program GWAS, RGS5 was found to be associated with lipoprotein disorders (PheCode 277.51)[97]. A locus was found on chromosome 21, gene NCAM2 (rs17794576, weight ranking 10). The protein encoded by this gene belongs to the immunoglobulin superfamily. It is a type I membrane protein and may function in selective fasciculation and zone-to-zone projection of the primary olfactory axons. Even though the revealed men/women AD subtype clusters do not share genes, they share several phenotypes such as height, smoking initiation, and educational attainment. Our

data-driven subtype cluster analysis revealed a select set of genes previously implicated in AD and dementia. The findings specifically suggest that these genes are linked to a distinct subtype of AD, significantly distinguishing it from other types of AD.

AD subtype and phenotypes – body weight

An association with weight was observed in the identified women AD subtype but not in the men one. The revealed women's AD subtype cluster was associated with lower weight compared to the rest of the ADNI sample. These results confirm previous studies showing that clinically important weight loss occurs more frequently in patients with AD than in cognitively normal control subjects[60]. Along the same lines, a meta-analysis that included 19 studies with 589,649 participants who were followed for up to 42 years concluded that, compared to a healthy weight, being underweight in midlife was associated with a 39% higher risk of dementia[98]. More broadly, there is evidence of a complex relationship between weight and AD/dementia[99]. For example, obesity or being overweight during middle age (roughly 35-65) can increase the risk of developing dementia later in life[100], while late-life obesity might be associated with reduced dementia risk[101]. Particularly in women, one of the factors that has been suggested to influence the association between body weight and dementia/AD is hormonal changes that occur during menopause; specifically, estrogen depletion has been found to be associated with obesity and the risk of AD[102]. Several factors were suggested for weight loss in people suffering from dementia/AD[103], such as reduced energy intake owing to their decreased mental status[104], sensory changes such as deterioration of the olfactory bulb and/or diminished gustatory perception as a result of cholinergic deficits[105], or dysphagia, or difficulties swallowing, may reduce energy intake[106]. AD-related increase[107] or decrease[108] body weight was also found to be associated with hypermetabolism, defined as an elevated basal metabolic rate of >10% or neuroendocrine dysregulation[109]. Our theoretical model proposes that what was once considered a single entity (AD diagnosis) is actually a collection of distinct subtypes, each defined by unique physiological and environmental variables. Our findings indicate that the identified female AD subtype is associated with lower body weight, along with distinct MRI biomarkers, and specific genetic variants (SNPs/genes). The revealed AD men subtype was not found to be associated with weight changes. Achieving a deeper understanding of each AD subtype requires integrating information from multiple modalities to distinguish between them effectively.

The SUMMER pipeline enables modular integration of different modalities and algorithms, optimizing the identification of AD subtypes. Relying on a single algorithm or combining all methods into one comprehensive model may obscure important distinctions, such as between different age groups or due to comparability that may eliminate key variables from the model. Integrating unsupervised algorithms into the analysis protocol is particularly beneficial for subtyping research, as these methods can explore large datasets in a blinded manner, independent of clinical diagnoses or self-reported information, and identify homogeneous clusters that help distinguish between different types of disorders. In this study, we identified distinct subtypes of men and women AD, each characterized by unique SNPs profiles and phenotypic characteristics. Recognition of these subtypes opens the door to earlier detection and the potential for personalized treatment strategies.

Limitations and future directions

We present a pipeline applied to the ADNI dataset, yielding promising results. However, the identified AD subtype may be influenced by the small sample size. Future research with larger datasets should reevaluate the generalizability of these findings and validate them. Additionally, further studies should be conducted to identify distinct subtypes that can be clearly differentiated from the ones we uncovered. These subtypes are likely to be associated with a different set of genes and a different set of brain structures. Furthermore, a comprehensive investigation is required to clarify the relationship between AD subtypes and different phenotypes, which will allow for a more precise definition of each subtype. One key research decision was to apply the pipeline to a mixed dataset comprising both CN and AD participants rather than exclusively to AD cases. While both approaches are valid, we chose the mixed dataset to demonstrate the model's applicability to individuals before establishing an AD diagnosis. Future research applying the SUMMER Pipeline to an AD-only sample would deepen understanding of AD subtypes.

Acknowledgments

Sivan Kinreich discloses support for the research from NIAAA AA029448.

References:

1 Sperling, R. A. *et al.* Association of Factors With Elevated Amyloid Burden in Clinically Normal Older Individuals. *JAMA Neurology* **77**, 735-745, doi:10.1001/jamaneurol.2020.0387 (2020).

2 Sperling, R. A. *et al.* The impact of amyloid-beta and tau on prospective cognitive decline in older individuals. *Annals of Neurology* **85**, 181-193, doi:https://doi.org/10.1002/ana.25395 (2019).

3 Bennett, D. A. *et al.* Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198-205, doi:10.1212/WNL.59.2.198 (2002).

4 Dokholyan, N. V., Mohs, R. C. & Bateman, R. J. Challenges and progress in research, diagnostics, and therapeutics in Alzheimer's disease and related dementias. *Alzheimer's & Dementia: Translational Research & Clinical Interventions* **8**, e12330, doi:https://doi.org/10.1002/trc2.12330 (2022).

5 Association, A. s. 2022 Alzheimer's disease facts and figures. *Alzheimers Dement* **18**, 700-789, doi:10.1002/alz.12638 (2022).

6 Brookmeyer, R., Abdalla, N., Kawas, C. H. & Corrada, M. M. Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the United States. *Alzheimers Dement* **14**, 121-129, doi:10.1016/j.jalz.2017.10.009 (2018).

7 Murray, M. E. *et al.* Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* **10**, 785-796, doi:10.1016/S1474-4422(11)70156-9 (2011).

8 Ferreira, D., Nordberg, A. & Westman, E. Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology* **94**, 436-448, doi:10.1212/WNL.0000000000009058 (2020).

9 Thal, D. R., Ghebremedhin, E., Orantes, M. & Wiestler, O. D. Vascular pathology in Alzheimer disease: correlation of cerebral amyloid angiopathy and arteriosclerosis/lipohyalinosis with cognitive decline. *Journal of neuropathology and experimental neurology* **62**, 1287-1301, doi:10.1093/jnen/62.12.1287 (2003).

10 Whitwell, J. L. *et al.* Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: a case-control study. *Lancet Neurol* **11**, 868-877, doi:10.1016/S1474-4422(12)70200-4 (2012).

11 Ferreira, D. *et al.* The contribution of small vessel disease to subtypes of Alzheimer's disease: a study on cerebrospinal fluid and imaging biomarkers. *Neurobiol Aging* **70**, 18-29, doi:10.1016/j.neurobiolaging.2018.05.028 (2018).

12 Murray, M. E. *et al.* Differential clinicopathologic and genetic features of late-onset amnestic dementias. *Acta Neuropathol* **128**, 411-421, doi:10.1007/s00401-014-1302-2 (2014).

13 Risacher, S. L. *et al.* Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline. *Neurology* **89**, 2176-2186, doi:10.1212/wnl.0000000000004670 (2017).

14 Corder, E. H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921-923, doi:10.1126/science.8346443 (1993).

15 Vogel, J. W. & Hansson, O. Subtypes of Alzheimer's disease: questions, controversy, and meaning. *Trends Neurosci* **45**, 342-345, doi:10.1016/j.tins.2022.02.001 (2022).

16 Cole, J. H., Marioni, R. E., Harris, S. E. & Deary, I. J. Brain age and other bodily 'ages': implications for neuropsychiatry. *Mol Psychiatry* **24**, 266-281, doi:10.1038/s41380-018-0098-1 (2019).

17 Le, Q. V., Zou, W. Y., Yeung, S. Y. & Ng, A. Y. in *CVPR 2011.* 3361-3368.

18 Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504-507, doi:10.1126/science.1127647 (2006).

19 Bengio, Y., Courville, A. C. & Vincent, P. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR, abs/1206.5538* **1**, 2012 (2012).
20 Kinreich, S. *et al.* Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: A Machine Learning study. *Molecular Psychiatry (In press)* (2019).
21 Whelan, R. *et al.* Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* **512**, 185-189, doi:10.1038/nature13402 (2014).
22 Petersen, R. C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology* **74**, 201-209, doi:10.1212/WNL.0b013e3181cb3e25 (2010).
23 Jack, C. R., Jr. *et al.* The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* **27**, 685-691, doi:10.1002/jmri.21049 (2008).
24 Petersen, R. C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology* **74**, 201-209, doi:10.1212/WNL.0b013e3181cb3e25 (2010).
25 Fischl, B. FreeSurfer. *Neuroimage* **62**, 774-781, doi:10.1016/j.neuroimage.2012.01.021 (2012).
26 Iwatsubo, T. [Alzheimer's disease Neuroimaging Initiative (ADNI)]. *Nihon Rinsho* **69 Suppl 8**, 570-574 (2011).
27 Risacher, S. L. *et al.* Longitudinal MRI atrophy biomarkers: Relationship to conversion in the ADNI cohort. *Neurobiology of Aging* **31**, 1401-1418, doi:https://doi.org/10.1016/j.neurobiolaging.2010.04.029 (2010).
28 Zhang, Y., Brady, M. & Smith, S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging* **20**, 45-57, doi:10.1109/42.906424 (2001).
29 Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54**, 2033-2044, doi:https://doi.org/10.1016/j.neuroimage.2010.09.025 (2011).
30 Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, doi:10.1016/j.neuroimage.2006.01.021 (2006).
31 Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23**, S208-S219, doi:https://doi.org/10.1016/j.neuroimage.2004.07.051 (2004).
32 Victoroff, J., Mack, W. J., Grafton, S. T., Schreiber, S. S. & Chui, H. C. A method to improve interrater reliability of visual inspection of brain MRI scans in dementia. *Neurology* **44**, 2267-2276, doi:10.1212/wnl.44.12.2267 (1994).
33 Sachdev, P., Cathcart, S., Shnier, R., Wen, W. & Brodaty, H. Reliability and validity of ratings of signal hyperintensities on MRI by visual inspection and computerised measurement. *Psychiatry Res* **92**, 103-115, doi:10.1016/s0925-4927(99)00036-0 (1999).
34 Wang, Z. *et al.* Idiosyncratic biases in the perception of medical images. *Front Psychol* **13**, 1049831, doi:10.3389/fpsyg.2022.1049831 (2022).
35 Wang, S. *et al.* AI-based MRI auto-segmentation of brain tumor in rodents, a multicenter study. *Acta neuropathologica communications* **11**, 11, doi:10.1186/s40478-023-01509-w (2023).
36 Lopez-Garcia, P., Aizenstein, H. J., Snitz, B. E., Walter, R. P. & Carter, C. S. Automated ROI-based brain parcellation analysis of frontal and temporal brain volumes in schizophrenia. *Psychiatry Res* **147**, 153-161, doi:10.1016/j.pscychresns.2006.04.007 (2006).
37 Cui, J. *et al.* Automatic segmentation of hemispheric CSF on MRI using deep learning: Quantifying cerebral edema following large hemispheric infarction. *Heliyon* **10**, e26673, doi:https://doi.org/10.1016/j.heliyon.2024.e26673 (2024).

38 Ardekani, B. A., Bermudez, E., Mubeen, A. M. & Bachman, A. H. Prediction of Incipient Alzheimer's Disease Dementia in Patients with Mild Cognitive Impairment. *J Alzheimers Dis* **55**, 269-281, doi:10.3233/jad-160594 (2017).
39 Ardekani, B. A., Hadid, S. A., Blessing, E. & Bachman, A. H. Sexual Dimorphism and Hemispheric Asymmetry of Hippocampal Volumetric Integrity in Normal Aging and Alzheimer Disease. *AJNR. American journal of neuroradiology* **40**, 276-282, doi:10.3174/ajnr.A5943 (2019).
40 Ardekani, B. A., Izadi, N. O., Hadid, S. A., Meftah, A. M. & Bachman, A. H. Effects of sex, age, and apolipoprotein E genotype on hippocampal parenchymal fraction in cognitively normal older adults. *Psychiatry Res Neuroimaging* **301**, 111107, doi:10.1016/j.pscychresns.2020.111107 (2020).
41 Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* **49**, 764-766, doi:https://doi.org/10.1016/j.jesp.2013.03.013 (2013).
42 Kong, X. Z. *et al.* Mapping brain asymmetry in health and disease through the ENIGMA consortium. *Hum Brain Mapp* **43**, 167-181, doi:10.1002/hbm.25033 (2022).
43 Liu, R. & Gillies, D. F. Overfitting in linear feature extraction for classification of high-dimensional image data. *Pattern Recognition* **53**, 73-86, doi:https://doi.org/10.1016/j.patcog.2015.11.015 (2016).
44 Saykin, A. J. *et al.* Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimers Dement* **6**, 265-273 (2010).
45 Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267-288 (1996).
46 Knight, K. & Fu, W. Asymptotics for lasso-type estimators. **28**, 1356-1378, doi:10.1214/aos/1015957397 (2000).
47 Shevade, S. K. & Keerthi, S. S. A simple and efficient algorithm for gene selection using sparse logistic regression. *Bioinformatics* **19**, 2246-2253 (2003).
48 Ahn, W. Y., Ramesh, D., Moeller, F. G. & Vassileva, J. Utility of Machine-Learning Approaches to Identify Behavioral Markers for Substance Use Disorders: Impulsivity Dimensions as Predictors of Current Cocaine Dependence. *Front Psychiatry* **7**, 34, doi:10.3389/fpsyt.2016.00034 (2016).
49 Moon, S. W. *et al.* Brain structure and allelic associations in Alzheimer's disease. *CNS neuroscience & therapeutics* **29**, 1034-1048, doi:10.1111/cns.14073 (2023).
50 Moezzi, B. *et al.* Characterization of Young and Old Adult Brains: An EEG Functional Connectivity Analysis. *Neuroscience* **422**, 230-239, doi:10.1016/j.neuroscience.2019.08.038 (2019).
51 Dibaji, M., Ospel, J., Souza, R. & Bento, M. Sex differences in brain MRI using deep learning toward fairer healthcare outcomes. *Front Comput Neurosci* **18**, 1452457, doi:10.3389/fncom.2024.1452457 (2024).
52 Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S. & Trapnell, C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nature Communications* **11**, 1537, doi:10.1038/s41467-020-15351-4 (2020).
53 Stolarek, I., Samelak-Czajka, A., Figlerowicz, M. & Jackowiak, P. Dimensionality reduction by UMAP for visualizing and aiding in classification of imaging flow cytometry data. *iScience* **25**, 105142, doi:https://doi.org/10.1016/j.isci.2022.105142 (2022).
54 Benjamini, Y. & Hochberg, Y. Vol. 57 289-300 (1995).
55 Saykin, A. J. *et al.* Genetic studies of quantitative MCI and AD phenotypes in ADNI: Progress, opportunities, and plans. *Alzheimers Dement* **11**, 792-814, doi:10.1016/j.jalz.2015.05.009 (2015).

56 Zhang, P., Jia, Y. & Shang, Y. Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks* **18**, 15501329221106935, doi:10.1177/15501329221106935 (2022).
57 Sagi, O. & Rokach, L. Approximating XGBoost with an interpretable decision tree. *Information Sciences* **572**, 522-542, doi:https://doi.org/10.1016/j.ins.2021.05.055 (2021).
58 Kang, S. Y. *et al.* Body mass index trajectories and the risk for Alzheimer's disease among older adults. *Scientific Reports* **11**, 3087, doi:10.1038/s41598-021-82593-7 (2021).
59 Hogan, D. B., Ebly, E. M. & Rockwood, K. Weight, blood pressure, osmolarity, and glucose levels across various stages of Alzheimer's disease and vascular dementia. *Dement Geriatr Cogn Disord* **8**, 147-151, doi:10.1159/000106623 (1997).
60 White, H., Pieper, C., Schmader, K. & Fillenbaum, G. Weight change in Alzheimer's disease. *J Am Geriatr Soc* **44**, 265-272, doi:10.1111/j.1532-5415.1996.tb00912.x (1996).
61 Kinreich, S. *et al.* Predicting alcohol use disorder remission: a longitudinal multimodal multi-featured machine learning approach. *Transl Psychiatry* **11**, 166, doi:10.1038/s41398-021-01281-2 (2021).
62 Schriml, L. M. *et al.* Modeling the enigma of complex disease etiology. *Journal of Translational Medicine* **21**, 148, doi:10.1186/s12967-023-03987-x (2023).
63 Rao, Y. L. *et al.* Hippocampus and its involvement in Alzheimer's disease: a review. *3 Biotech* **12**, 55, doi:10.1007/s13205-022-03123-4 (2022).
64 Bird, C. M. & Burgess, N. The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience* **9**, 182-194, doi:10.1038/nrn2335 (2008).
65 Boutet, C. *et al.* Detection of volume loss in hippocampal layers in Alzheimer's disease using 7 T MRI: A feasibility study. *NeuroImage: Clinical* **5**, 341-348, doi:https://doi.org/10.1016/j.nicl.2014.07.011 (2014).
66 Moreno-Jiménez, E. P. *et al.* Adult hippocampal neurogenesis is abundant in neurologically healthy subjects and drops sharply in patients with Alzheimer's disease. *Nature Medicine* **25**, 554-560, doi:10.1038/s41591-019-0375-9 (2019).
67 Allen, G. *et al.* Reduced Hippocampal Functional Connectivity in Alzheimer Disease. *Archives of Neurology* **64**, 1482-1487, doi:10.1001/archneur.64.10.1482 (2007).
68 Bogousslavsky, J., Miklossy, J., Deruaz, J. P., Assal, G. & Regli, F. Lingual and fusiform gyri in visual processing: a clinico-pathologic study of superior altitudinal hemianopia. *J Neurol Neurosurg Psychiatry* **50**, 607-614, doi:10.1136/jnnp.50.5.607 (1987).
69 Van Hoesen, G. W., Augustinack, J. C., Dierking, J., Redman, S. J. & Thangavel, R. The parahippocampal gyrus in Alzheimer's disease. Clinical and preclinical neuroanatomical correlates. *Annals of the New York Academy of Sciences* **911**, 254-274, doi:10.1111/j.1749-6632.2000.tb06731.x (2000).
70 Whitwell, J. L. *et al.* MRI patterns of atrophy associated with progression to AD in amnestic mild cognitive impairment. *Neurology* **70**, 512-520, doi:10.1212/01.wnl.0000280575.77437.a2 (2008).
71 Ma, D. *et al.* The fusiform gyrus exhibits an epigenetic signature for Alzheimer's disease. *Clinical Epigenetics* **12**, 129, doi:10.1186/s13148-020-00916-3 (2020).
72 Kawagoe, T., Matsushita, M., Hashimoto, M., Ikeda, M. & Sekiyama, K. Face-specific memory deficits and changes in eye scanning patterns among patients with amnestic mild cognitive impairment. *Scientific Reports* **7**, 14344, doi:10.1038/s41598-017-14585-5 (2017).
73 Robinson, J. L. *et al.* The functional connectivity of the human caudate: an application of meta-analytic connectivity modeling with behavioral filtering. *Neuroimage* **60**, 117-129, doi:10.1016/j.neuroimage.2011.12.010 (2012).

74 de Jong, L. W. *et al.* Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain* **131**, 3277-3285, doi:10.1093/brain/awn278 (2008).
75 Kim, S. J. *et al.* Development of prediction models for distinguishable cognitive trajectories in patients with amyloid positive mild cognitive impairment. *Neurobiology of Aging* **114**, 84-93, doi:https://doi.org/10.1016/j.neurobiolaging.2022.02.012 (2022).
76 Johnson, J. C. S. *et al.* Hearing and dementia: from ears to brain. *Brain* **144**, 391-401, doi:10.1093/brain/awaa429 (2021).
77 Hänggi, J., Streffer, J., Jäncke, L. & Hock, C. Volumes of lateral temporal and parietal structures distinguish between healthy aging, mild cognitive impairment, and Alzheimer's disease. *J Alzheimers Dis* **26**, 719-734, doi:10.3233/jad-2011-101260 (2011).
78 Jacobs, H. I., Van Boxtel, M. P., Jolles, J., Verhey, F. R. & Uylings, H. B. Parietal cortex matters in Alzheimer's disease: an overview of structural, functional and metabolic findings. *Neurosci Biobehav Rev* **36**, 297-309, doi:10.1016/j.neubiorev.2011.06.009 (2012).
79 Cattaneo, L., Maule, F., Tabarelli, D., Brochier, T. & Barchiesi, G. Online repetitive transcranial magnetic stimulation (TMS) to the parietal operculum disrupts haptic memory for grasping. *Hum Brain Mapp* **36**, 4262-4271, doi:10.1002/hbm.22915 (2015).
80 Sepulcre, J. Functional streams and cortical integration in the human brain. *Neuroscientist* **20**, 499-508, doi:10.1177/1073858414531657 (2014).
81 Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704-712, doi:10.1038/s41586-022-05275-y (2022).
82 Keaton, J. M. *et al.* Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nature Genetics* **56**, 778-791, doi:10.1038/s41588-024-01714-w (2024).
83 Pirruccello, J. P. *et al.* Deep learning enables genetic analysis of the human thoracic aorta. *Nature Genetics* **54**, 40-51, doi:10.1038/s41588-021-00962-4 (2022).
84 Koskeridis, F. *et al.* Multi-trait association analysis reveals shared genetic loci between Alzheimer's disease and cardiovascular traits. *Nat Commun* **15**, 9827, doi:10.1038/s41467-024-53452-6 (2024).
85 Wang, H. *et al.* Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. *Neurobiology of Aging* **93**, 61-68, doi:https://doi.org/10.1016/j.neurobiolaging.2020.04.025 (2020).
86 Luna-Viramontes, N. I. *et al.* PHF-Core Tau as the Potential Initiating Event for Tau Pathology in Alzheimer's Disease. *Frontiers in cellular neuroscience* **14**, 247, doi:10.3389/fncel.2020.00247 (2020).
87 Zhang, Q. *et al.* Genotype effects contribute to variation in longitudinal methylome patterns in older people. *Genome medicine* **10**, 75, doi:10.1186/s13073-018-0585-7 (2018).
88 van der Weyden, L. & Adams, D. J. The Ras-association domain family (RASSF) members and their role in human tumourigenesis. *Biochimica et biophysica acta* **1776**, 58-85, doi:10.1016/j.bbcan.2007.06.003 (2007).
89 Suhre, K. Genetic associations with ratios between protein levels detect new pQTLs and reveal protein-protein interactions. *Cell genomics* **4**, 100506, doi:10.1016/j.xgen.2024.100506 (2024).
90 Western, D. *et al.* Proteogenomic analysis of human cerebrospinal fluid identifies neurologically relevant regulation and implicates causal proteins for Alzheimer’s disease. *Nature Genetics* **56**, 2672-2684, doi:10.1038/s41588-024-01972-8 (2024).
91 van der Meer, D. *et al.* Boosting Schizophrenia Genetics by Utilizing Genetic Overlap With Brain Morphology. *Biological Psychiatry* **92**, 291-298, doi:https://doi.org/10.1016/j.biopsych.2021.12.007 (2022).

92 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, doi:10.1038/nature17671 (2016).
93 Babb, S. G. *et al.* Zebrafish R-cadherin (Cdh4) controls visual system development and differentiation. *Developmental dynamics : an official publication of the American Association of Anatomists* **233**, 930-945, doi:10.1002/dvdy.20431 (2005).
94 Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics* **54**, 437-449, doi:10.1038/s41588-022-01016-z (2022).
95 Saunders, G. R. B. *et al.* Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720-724, doi:10.1038/s41586-022-05477-4 (2022).
96 Ruth, K. S. *et al.* Using human genetics to understand the disease impacts of testosterone in men and women. *Nature Medicine* **26**, 252-258, doi:10.1038/s41591-020-0751-5 (2020).
97 Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* **385**, eadj1182, doi:doi:10.1126/science.adj1182 (2024).
98 Albanese, E. *et al.* Body mass index in midlife and dementia: Systematic review and meta-regression analysis of 589,649 men and women followed in longitudinal studies. *Alzheimers Dement (Amst)* **8**, 165-178, doi:10.1016/j.dadm.2017.05.007 (2017).
99 Kang, S. Y. *et al.* Body mass index trajectories and the risk for Alzheimer's disease among older adults. *Sci Rep* **11**, 3087, doi:10.1038/s41598-021-82593-7 (2021).
100 Whitmer, R. A., Gunderson, E. P., Barrett-Connor, E., Quesenberry, C. P., Jr. & Yaffe, K. Obesity in middle age and future risk of dementia: a 27 year longitudinal population based study. *Bmj* **330**, 1360, doi:10.1136/bmj.38446.466238.E0 (2005).
101 Sun, Z. *et al.* Late-life obesity is a protective factor for prodromal Alzheimer's disease: a longitudinal study. *Aging (Albany NY)* **12**, 2005-2017, doi:10.18632/aging.102738 (2020).
102 Christensen, A. & Pike, C. J. Menopause, obesity and inflammation: interactive risk factors for Alzheimer's disease. *Front Aging Neurosci* **7**, 130, doi:10.3389/fnagi.2015.00130 (2015).
103 Franx, B. A. A., Arnoldussen, I. A. C., Kiliaan, A. J. & Gustafson, D. R. Weight Loss in Patients with Dementia: Considering the Potential Impact of Pharmacotherapy. *Drugs & aging* **34**, 425-436, doi:10.1007/s40266-017-0462-x (2017).
104 Miyamoto, K., Higashino, S., Mochizuki, K., Goda, T. & Koyama, H. Evaluation of weight loss in the community-dwelling elderly with dementia as assessed by eating behavior and mental status. *Asia Pacific journal of clinical nutrition* **20**, 9-13 (2011).
105 Aziz, N. A. *et al.* Weight loss in neurodegenerative disorders. *J Neurol* **255**, 1872-1880, doi:10.1007/s00415-009-0062-8 (2008).
106 Easterling, C. S. & Robbins, E. Dementia and dysphagia. *Geriatric nursing (New York, N.Y.)* **29**, 275-285, doi:10.1016/j.gerinurse.2007.10.015 (2008).
107 Wolf-Klein, G. P. *et al.* Energy requirements in Alzheimer's disease patients. *Nutrition (Burbank, Los Angeles County, Calif.)* **11**, 264-268 (1995).
108 Poehlman, E. T. *et al.* Daily energy expenditure in free-living non-institutionalized Alzheimer's patients: a doubly labeled water study. *Neurology* **48**, 997-1002, doi:10.1212/wnl.48.4.997 (1997).
109 Polleri, A., Gianelli, M. V. & Murialdo, G. Dementia: a neuroendocrine perspective. *Journal of endocrinological investigation* **25**, 73-83, doi:10.1007/bf03343964 (2002).