



## OPEN ACCESS

## EDITED BY

Gal Bitan,  
David Geffen School of Medicine (UC), Los Angeles, United States

## REVIEWED BY

Marcia H. Ratner,  
Boston University School of Medicine,  
United States  
Hao Wang,  
Shanghai Jiao Tong University, China

## \*CORRESPONDENCE

Alejo J. Nevado-Holgado  
alejo.nevado-holgado@psych.ox.ac.uk  
Liu Shi  
shiliuswgch@gmail.com

## SPECIALTY SECTION

This article was submitted to  
Cellular and Molecular Mechanisms of  
Brain-aging,  
a section of the journal  
Frontiers in Aging Neuroscience

RECEIVED 08 September 2022

ACCEPTED 04 November 2022

PUBLISHED 29 November 2022

## CITATION

Zhang Y, Ghose U, Buckley NJ,  
Engelborghs S, Sleegers K, Frisoni GB,  
Wallin A, Lleó A, Popp J, Martinez-Lage P,  
Legido-Quigley C, Barkhof F, Zetterberg H,  
Visser PJ, Bertram L, Lovestone S,  
Nevado-Holgado AJ and Shi L (2022)  
Predicting AT(N) pathologies in Alzheimer's  
disease from blood-based proteomic data  
using neural networks.  
*Front. Aging Neurosci.* 14:1040001.  
doi: 10.3389/fnagi.2022.1040001

## COPYRIGHT

© 2022 Zhang, Ghose, Buckley,  
Engelborghs, Sleegers, Frisoni, Wallin, Lleó,  
Popp, Martinez-Lage, Legido-Quigley,  
Barkhof, Zetterberg, Visser, Bertram,  
Lovestone, Nevado-Holgado and Shi. This  
is an open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Predicting AT(N) pathologies in Alzheimer's disease from blood-based proteomic data using neural networks

Yuting Zhang<sup>1</sup>, Upamanyu Ghose<sup>1</sup>, Noel J. Buckley<sup>1</sup>,  
Sebastian Engelborghs<sup>2,3,4</sup>, Kristel Sleegers<sup>5,6</sup>,  
Giovanni B. Frisoni<sup>7</sup>, Anders Wallin<sup>8</sup>, Alberto Lleó<sup>9</sup>, Julius  
Popp<sup>10,11</sup>, Pablo Martinez-Lage<sup>12</sup>, Cristina Legido-Quigley<sup>13,14</sup>,  
Frederik Barkhof<sup>15,16</sup>, Henrik Zetterberg<sup>8,17,18,19,20</sup>, Pieter Jelle  
Visser<sup>21,22</sup>, Lars Bertram<sup>23,24</sup>, Simon Lovestone<sup>1,25</sup>,  
Alejo J. Nevado-Holgado<sup>1\*</sup> and Liu Shi<sup>1\*</sup>

<sup>1</sup>Department of Psychiatry, University of Oxford, Oxford, United Kingdom, <sup>2</sup>Department of Biomedical Sciences, Reference Center for Biological Markers of Dementia (BIODEM), Institute Born-Bunge, University of Antwerp, Antwerp, Belgium, <sup>3</sup>Department of Neurology, Universitair Ziekenhuis Brussel, Brussels, Belgium, <sup>4</sup>Center for Neurosciences (C4N), Vrije Universiteit Brussel, Brussels, Belgium, <sup>5</sup>Complex Genetics Group, VIB Center for Molecular Neurology, Antwerp, Belgium, <sup>6</sup>Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium, <sup>7</sup>Department of Psychiatry, University of Geneva, Geneva, Switzerland, <sup>8</sup>Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden, <sup>9</sup>Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, <sup>10</sup>Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland, <sup>11</sup>Department of Geriatric Psychiatry, University Hospital of Psychiatry and University of Zürich, Zürich, Switzerland, <sup>12</sup>CITA-Alzheimer Foundation, San Sebastian, Spain, <sup>13</sup>Kings College London, London, United Kingdom, <sup>14</sup>The Systems Medicine Group, Steno Diabetes Center, Gentofte, Denmark, <sup>15</sup>Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, Netherlands, <sup>16</sup>University College London (UCL) Institutes of Neurology and Healthcare Engineering, London, United Kingdom, <sup>17</sup>Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden, <sup>18</sup>UK Dementia Research Institute at UCL, London, United Kingdom, <sup>19</sup>Department of Neurodegenerative Disease, UCL Institute of Neurology, London, United Kingdom, <sup>20</sup>Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China, <sup>21</sup>Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Alzheimer Centrum Limburg, Maastricht University, Maastricht, Netherlands, <sup>22</sup>Alzheimer Center, VU University Medical Center, Amsterdam, Netherlands, <sup>23</sup>Lübeck Interdisciplinary Platform for Genome Analytics, University of Lübeck, Lübeck, Germany, <sup>24</sup>Department of Psychology, University of Oslo, Oslo, Norway, <sup>25</sup>Janssen R&D, High Wycombe, United Kingdom

**Background and objective:** Blood-based biomarkers represent a promising approach to help identify early Alzheimer's disease (AD). Previous research has applied traditional machine learning (ML) to analyze plasma omics data and search for potential biomarkers, but the most modern ML methods based on deep learning has however been scarcely explored. In the current study, we aim to harness the power of state-of-the-art deep learning neural networks (NNs) to identify plasma proteins that predict amyloid, tau, and neurodegeneration (AT[N]) pathologies in AD.

**Methods:** We measured 3,635 proteins using SOMAscan in 881 participants from the European Medical Information Framework for AD Multimodal Biomarker

Discovery study (EMIF-AD MBD). Participants underwent measurements of brain amyloid  $\beta$  ( $A\beta$ ) burden, phosphorylated tau (p-tau) burden, and total tau (t-tau) burden to determine their AT(N) statuses. We ranked proteins by their association with  $A\beta$ , p-tau, t-tau, and AT(N), and fed the top 100 proteins along with age and apolipoprotein E (*APOE*) status into NN classifiers as input features to predict these four outcomes relevant to AD. We compared NN performance of using proteins, age, and *APOE* genotype with performance of using age and *APOE* status alone to identify protein panels that optimally improved the prediction over these main risk factors. Proteins that improved the prediction for each outcome were aggregated and nominated for pathway enrichment and protein–protein interaction enrichment analysis.

**Results:** Age and *APOE* alone predicted  $A\beta$ , p-tau, t-tau, and AT(N) burden with area under the curve (AUC) scores of 0.748, 0.662, 0.710, and 0.795. The addition of proteins significantly improved AUCs to 0.782, 0.674, 0.734, and 0.831, respectively. The identified proteins were enriched in five clusters of AD-associated pathways including human immunodeficiency virus 1 infection, p53 signaling pathway, and phosphoinositide-3-kinase–protein kinase B/Akt signaling pathway.

**Conclusion:** Combined with age and *APOE* genotype, the proteins identified have the potential to serve as blood-based biomarkers for AD and await validation in future studies. While the NNs did not achieve better scores than the support vector machine model used in our previous study, their performances were likely limited by small sample size.

#### KEYWORDS

Alzheimer's disease, plasma proteomics, amyloid  $\beta$ , tau, neurodegeneration, machine learning, artificial neural networks

## Introduction

Alzheimer's disease (AD) is a growing public health concern with no disease-modifying treatment available (Apostolova, 2016). The core criteria for clinical diagnosis of AD are based on behavioral and cognitive symptoms, but neuropathological changes in the central nervous system initiate years before the onset of cognitive impairment (Jack et al., 2013). The preclinical stage when pathologies develop in the absence of clinical symptoms presents an opportunity for early intervention with drugs to slow down or even halt disease progression. Currently, the well-established biomarkers for AD are cerebrospinal fluid (CSF) amyloid- $\beta$  peptide ( $A\beta$ ) and amyloid positron emission tomography (PET). They both highly correlate with pathologies found in brain autopsy (Strozyk et al., 2003; Curtis et al., 2015) and can identify early AD with high accuracy (Palmqvist et al., 2015). However, collecting CSF requires a lumbar puncture, an invasive practice that could lead to complications including post-dural puncture headache (de Almeida et al., 2011). PET imaging is expensive and requires specialist equipment that is not easily available. Thus, these two approaches have limited clinical utility. Blood-based biomarkers offer a desirable strategy to aid early diagnosis of AD. As blood-based test is minimally invasive,

economical, and widely available, they can serve as efficient prescreening tools in a multi-stage diagnostic process (O'Bryant et al., 2016). Earlier research has provided evidence of alterations of proteomic profiles in blood samples associated with AD state (Hye et al., 2006; Ray et al., 2007), validating the feasibility of blood-based biomarkers.

The development of blood-based biomarkers is facilitated by advances in not only targeted approaches (e.g., plasma phosphorylated tau measurements; Ashton et al., 2021; Milà-Alomà et al., 2022) but also untargeted large-scale omics technologies. Researchers have adopted high-resolution mass spectrometry for proteomic profiling of blood and discovery of AD protein signatures (Lopez et al., 2005; Dey et al., 2019). Highly sensitive multiplexed immunoassay platforms, such as Olink (Whelan et al., 2019; Jiang et al., 2022), and aptamer-based assay platforms, such as SOMAscan (Kiddle et al., 2014; Sattler et al., 2014), have further allowed researchers to capture the complexity of plasma proteome and identify prospective biomarkers by measuring thousands of proteins simultaneously in thousands of individuals with a single platform. In analyzing the rich omics data, machine learning (ML), a subdomain of artificial intelligence, has proven an invaluable tool (Li et al., 2021). Previous studies have had great success in finding blood analytes that predict

AD-related measures using traditional ML algorithms such as support vector machines (Voyle et al., 2015; Shi L. et al., 2019; Karaglani et al., 2020), decision trees (Pérez-Grijalba et al., 2019; Shi Y. et al., 2019), and random forests (Goudey et al., 2019; Beltrán et al., 2020; Lin et al., 2020; Zhao et al., 2020).

Deep learning neural networks (NNs) are the last iteration of ML methods and have significant advantages when compared to the older ML methods both in terms of classification accuracy and versatility, yet fewer studies have explored their utility in blood-based AD biomarker discovery. As they are capable of learning data representations in multiple hierarchies, they have often outperformed conventional models in various applications including predicting clinical diagnoses (Durstewitz et al., 2019). Thus, in the current study, we implement deep NNs to analyze plasma proteomics measured by SOMAscan. Our objective is to identify candidate plasma protein panels to detect amyloid, tau, and neurodegeneration (AT[N]) pathologies in AD.

## Materials and methods

### Participants

European Medical Information Framework (EMIF, [www.emif.eu](http://www.emif.eu)) is funded by the Innovative Medicines Initiative to support the reuse of existing healthcare data. As part of this project, EMIF-AD set up the Multimodal Biomarker Discovery (MBD) study by integrating multi-omics data to facilitate the development of AD biomarkers. Participants of the current study were from the EMIF 1000 sub-cohort assembled in a previous study (Bos et al., 2018). The original sub-cohort included 1,221 participants from 11 single- or multi-center studies across Europe. In the current study, we included only subjects who had available plasma samples, resulting in a subset of 881 participants from 10 studies. Among them, 311 had normal cognition (CN), 386 had mild cognitive impairment (MCI), and 184 had a diagnosis of AD dementia.

All 881 participants underwent measurement of the concentration of the 42 amino acid-long  $A\beta$  protein,  $A\beta_{42}$ , in CSF, sometimes in a ratio with the 40 amino acid-long form,  $A\beta_{40}$ , or amyloid PET as the primary AD outcome. A previous study has classified the amyloid burden of each participant as high or low based on z-score cutoffs (Bos et al., 2018). In addition, a subgroup of 787 subjects underwent measurement of CSF phosphorylated tau (p-tau), and a subgroup of 791 subjects underwent measurement of CSF total tau (t-tau). CSF p-tau and t-tau levels were both measured locally, and their statuses were classified as high or low with local cutoffs. Mini-mental state examination (MMSE) was administered to a majority of participants to assess cognitive function. All participants were genotyped to determine whether they were apolipoprotein E (*APOE*) risk allele  $\epsilon 4$  carriers or non-carriers. Lastly, proteins in plasma samples were measured by the SOMAscan assay (SomaLogic Inc.) in a previous study (Shi L. et al., 2019).

SOMAscan is an aptamer-based platform which transforms individual protein signals into corresponding chemically-modified nucleotide signals that can be quantified by relative fluorescence on DNA microarrays (Gold et al., 2010). Plasma samples were divided into two groups, and each group was processed independently. Forty samples were tested in both batches to normalize the measurements across assay runs. A total of 3,635 plasma proteins were quantified.

### Statistical analysis

All statistical analyses were performed using Python (version 3.9.7). We compared the demographic and clinical characteristics of patients with different diagnoses. Continuous variables were compared between groups by the Kruskal-Wallis one-way ANOVA test followed by Mann-Whitney *U*-tests for pairwise comparisons. Categorical variables were compared between groups by Chi-square test.

### Building NNs

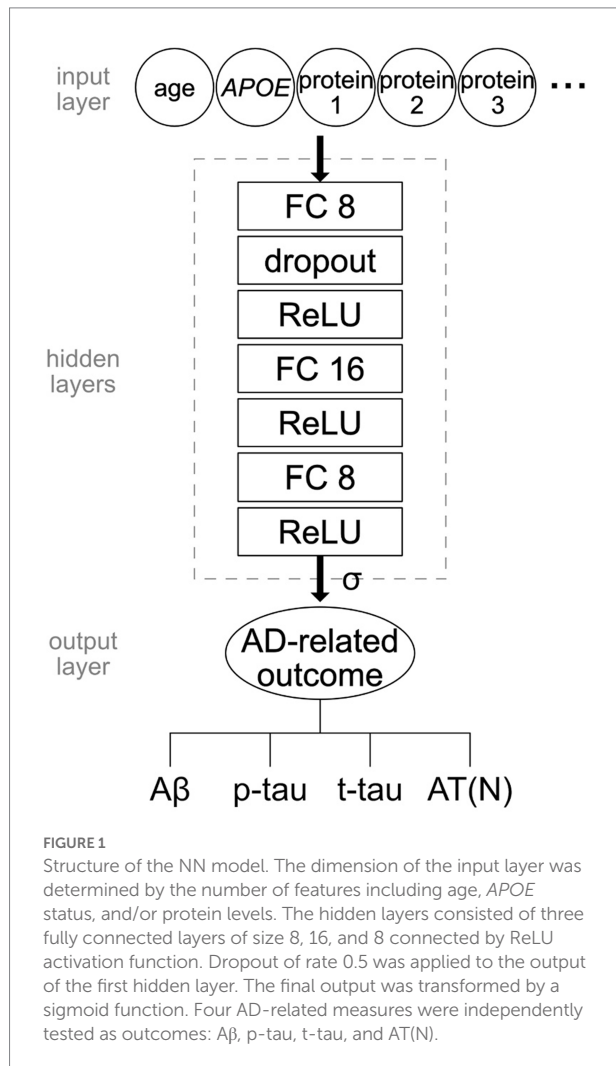
NNs were implemented using the Python package PyTorch (version 1.10.1). The network (Figure 1) consisted of three fully connected hidden layers with sizes 8, 16, and 8. Dimension of the input layer varied depending on the number of features: one for age, one for *APOE* status, and one for each protein. The output layer had one dimension to encode the binary outcome. Dropout was applied to the output of the first hidden layer at a rate of 0.5. Each hidden layer was followed by a rectified linear unit (ReLU) activation function. The loss function was binary cross-entropy with sigmoid. Learning was implemented using the Adam stochastic optimization algorithm with a learning rate of 0.01.

Training and testing were implemented with 5-fold cross-validation. The dataset was shuffled across samples before splitting between training and testing. On each training/testing split, the parameters of the NN (i.e., weights) were reinitialized, and the NN was trained for 15 epochs. In each epoch, data was fed into NN in minibatches of size 128. The outputs were transformed by a sigmoid function to compare with the binary ground-truth values.

### Discrimination of AT(N)

The dataset was preprocessed for analysis. Age was power-transformed and z-scored. All protein levels were power-transformed and z-scored to have zero mean and unit variance. The effects of study and blood freeze-thaw cycles on proteins were removed by linear regression, and the resulting residuals replaced raw protein levels.

We aim to predict  $A\beta$  burden (high vs. low), p-tau burden (high vs. low), t-tau burden (high vs. low), and AT(N) profile (high



A $\beta$ /p-tau/t-tau vs. low A $\beta$ /p-tau/t-tau) from plasma proteins along with age and *APOE* genotype. For each of these classification objectives, we first performed logistic regression analysis to measure the linear association between each protein and the target and ranked the proteins by ascending *p* values. We then selected from 1 to 100 top-ranked proteins along with age and *APOE* status as input features for NN classification and compared their performance with that of age and *APOE*  $\epsilon 4$  status alone. For each unique set of features and the target, we repeated training–testing 10 times to obtain the average receiver operating characteristic (ROC) curve and the area under the curve (AUC). AUC ROC scores resulting from different features were compared using independent *t*-tests.

## Enrichment analysis

The proteins that achieved the best performance in NN classification along with age and *APOE* status in differentiating high and low A $\beta$ , p-tau, t-tau, and AT(N) were aggregated and nominated for pathway enrichment and protein–protein

**TABLE 1** Demographic and clinical characteristics of the study population.

Characteristic	Sample size	CN	MCI	AD	<i>P</i> -value
<i>N</i>	881	311	386	184	
Age, median (IQR)	881	66.0 (58.9–70.0)	70.5 (65.1–75.5)	71.0 (63.4–77.0)	<0.001*
Male sex, <i>N</i> (%)	881	133 (42.8%)	181 (46.9%)	86 (46.7%)	0.509
MMSE, median (IQR)	878	29.0 (28.0–30.0)	26.5 (25.0–28.0)	22.0 (18.0–25.0)	<0.001*
<i>APOE</i> $\epsilon 4+$ , <i>N</i> (%)	881	118 (37.9%)	186 (48.2%)	112 (60.9%)	<0.001*
A $\beta$ +, <i>N</i> (%)	881	93 (30.0%)	245 (63.5%)	165 (89.7%)	<0.001*
P-tau+, <i>N</i> (%)	787	42 (19.1%)	203 (53.0%)	128 (69.6%)	<0.001*
T-tau+, <i>N</i> (%)	791	45 (20.1%)	221 (57.6%)	149 (81.4%)	<0.001*

CN, normal cognition; MCI, mild cognitive impairment; AD, Alzheimer’s disease; IQR, interquartile range; MMSE, mini-mental state examination; *APOE*, apolipoprotein E; A $\beta$ , amyloid- $\beta$ ; p-tau, phosphorylated-tau; t-tau, total tau. \**p* < 0.05.

interaction enrichment analysis. The analysis was performed using the Metascape software (Zhou et al., 2019). The complete set of proteins was provided as “background.” The inputs were searched against KEGG Pathway database for pathway enrichment analysis and STRING, BioGrid, OmniPath, and InWeb\_IM databases for protein–protein interaction analysis.

## Results

### Demographic and clinical variables

The current study included 881 participants from the EMIF 1000 sub-cohort. The demographic and clinical variables for each diagnostic group are described in Table 1. Patients with AD or MCI were older than CN subjects (AD vs. CN: odds ratio [OR] = 1.08, *p* < 0.001; MCI vs. CN: OR = 1.08, *p* < 0.001). CN subjects had higher MMSE scores than MCI subjects (OR = 2.08, *p* < 0.001), and MCI subjects had higher MMSE scores than AD subjects (OR = 1.44, *p* < 0.001). AD patients also had a higher prevalence of *APOE*  $\epsilon 4$  carriers (AD vs. MCI: OR = 1.67, *p* < 0.001; AD vs. CN: OR = 2.54, *p* < 0.01). There was no statistical difference in sex distribution between diagnostic groups (AD vs. MCI: OR = 1.01, *p* = 1.00; MCI vs. CN: OR = 0.85, *p* = 0.31; AD vs. CN: OR = 1.01, *p* = 0.44). For AT(N) biomarkers, AD subjects had a higher prevalence of low CSF A $\beta_{42}$  or A $\beta_{42/40}$  or positive amyloid PET, high CSF p-tau, and high CSF t-tau than MCI subjects (A $\beta$ : OR = 5.00; p-tau: OR = 2.03; t-tau: OR = 3.23; all *p* < 0.001) and CN subjects (A $\beta$ : OR = 20.36; p-tau: OR = 9.69; t-tau: OR = 17.43; all *p* < 0.001).

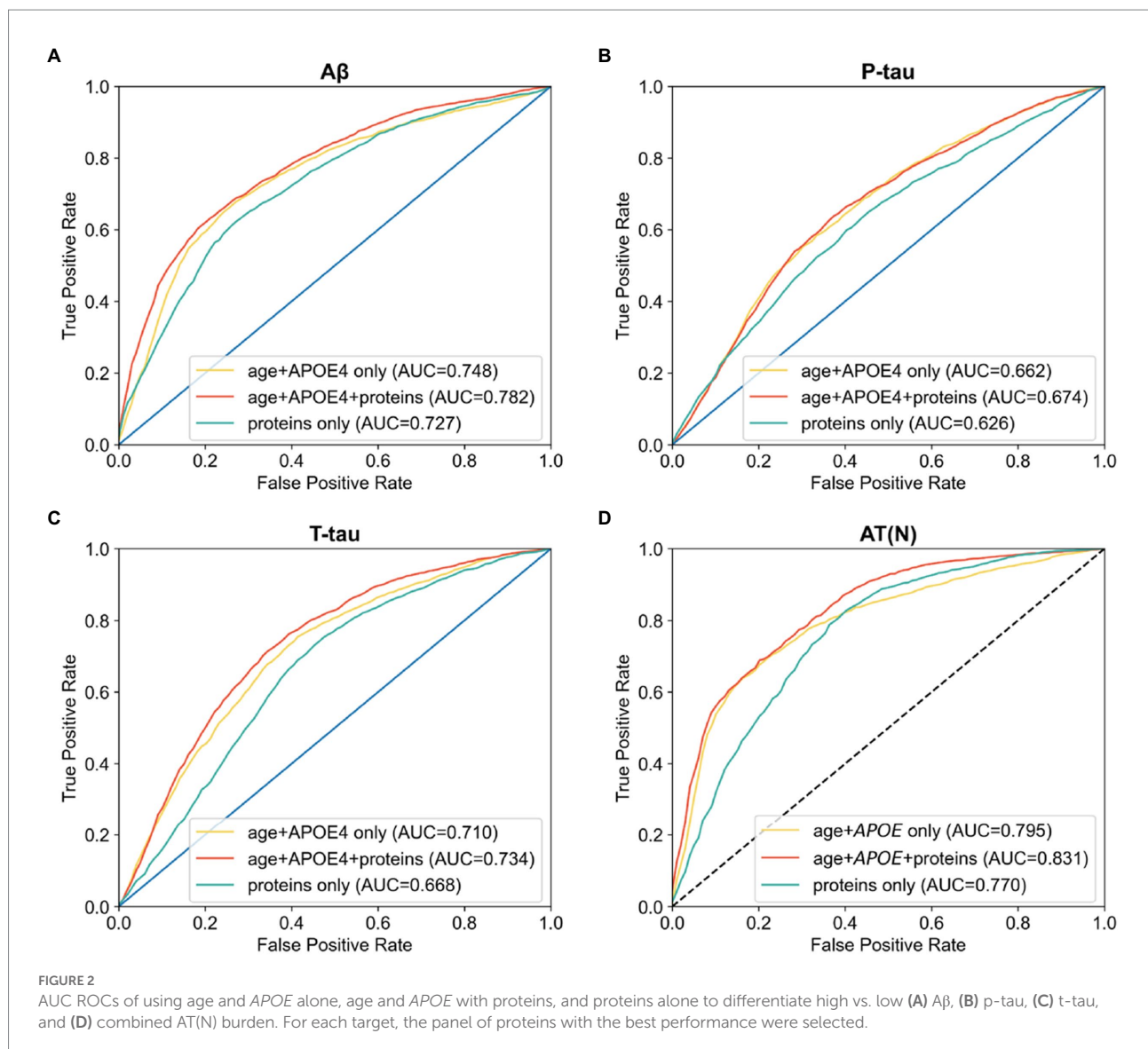
## Discrimination of AT(N) markers using NNS

We first used regression analysis to find the linear association between each protein and A $\beta$ . Binary A $\beta$  burden (high:  $N=503$ ; low:  $N=378$ ) was defined by z-score cutoff of CSF A $\beta_{42/40}$  at 0.061 or local cutoffs of CSF A $\beta_{42}$  and amyloid PET. Out of all proteins, 1,793 reached statistical significance (uncorrected  $p < 0.05$ ), and 1,492 of them reached the false discovery rate after correction for multiple comparisons (corrected  $p < 0.05$ ). We then sought to find the optimal set of proteins that differentiate high vs. low A $\beta$  burden using NNS. The combination of age and APOE  $\epsilon 4$  alone achieved an AUC of 0.748 (95% confidence interval [CI] 0.745–0.750). With protein features alone, a panel of 15 proteins achieved the highest AUC of 0.727 (95% CI 0.722–0.732). After the combination of age, APOE  $\epsilon 4$ , and proteins as input features, we found a panel of 11 proteins (see [Supplementary Table 1](#)) that

achieved the highest predictive value with an AUC of 0.782 (95% CI 0.779–0.785). This was significantly higher than AUCs from using only age and APOE  $\epsilon 4$  status ( $p < 0.001$ ; [Figure 2A](#)).

Binary p-tau burden (high:  $N=373$ ; low:  $N=414$ ) was defined by local cutoffs of CSF p-tau levels. For p-tau, 690 proteins reached statistical significance (uncorrected  $p < 0.05$ ) while none passed the false discovery rate (corrected  $p < 0.05$ ). In classifying high and low p-tau, age and APOE  $\epsilon 4$  alone achieved an AUC of 0.662 (95% CI 0.658–0.666). With protein features alone, the set of 99 proteins achieved the highest AUC of 0.626 (95% CI 0.615–0.636). After the combination of age, APOE  $\epsilon 4$ , and proteins, we found that the addition of 2 proteins (see [Supplementary Table 1](#)) achieved the highest AUC of 0.674 (95% CI 0.669–0.678), which was significantly greater than the AUCs from age and APOE alone ( $p < 0.001$ ; [Figure 2B](#)).

Binary t-tau burden (high:  $N=415$ ; low:  $N=376$ ) was defined by local cutoffs of CSF t-tau levels. For t-tau, 1,437 out



of all proteins reached statistical significance (uncorrected  $p < 0.05$ ), and 1,038 of them reached the false discovery rate (corrected  $p < 0.05$ ). Age and *APOE*  $\epsilon 4$  together achieved an AUC of 0.710 (95% CI 0.708–0.713). With protein features alone, a panel of 33 proteins achieved the highest AUC of 0.668 (95% CI 0.660–0.675). Combining with age and *APOE*  $\epsilon 4$ , a panel of 29 proteins (see [Supplementary Table 1](#)) achieved the highest AUC of 0.734 (95% CI 0.729–0.738), significantly exceeding that of age and *APOE* alone ( $p < 0.001$ ; [Figure 2C](#)).

## Discrimination of A + T + N+ from A – T – N – using NNs

We further used proteins to differentiate subjects with extreme AT(N) profiles using NNs. High and low AT(N) levels (A + T + N+:  $N = 298$ ; A – T – N–:  $N = 229$ ) were defined using the cutoffs of binary A $\beta$ , p-tau, and t-tau burden. In logistic regression analysis, 1,638 proteins reached statistical significance (uncorrected  $p < 0.05$ ) in association with A + T + N+ and A – T – N–. 1,317 of them passed the false discovery rate (corrected  $p < 0.05$ ). The correlations of protein  $p$  values between each pair of AD-related outcomes are listed in [Supplementary Table 2](#).

In differentiating A + T + N+ and A – T – N–, age and *APOE*  $\epsilon 4$  alone achieved an AUC of 0.795 (95% CI 0.793–0.798). Using the top 100 proteins as input features, we found that 86 proteins achieved the highest AUC of 0.770 (95% CI 0.765–0.774). The combination of age, *APOE*  $\epsilon 4$ , and proteins showed that a panel of 7 proteins (see [Supplementary Table 1](#)) reached the highest AUC of 0.831 (95% CI 0.827–0.835), which was significantly better than those obtained from age and *APOE*  $\epsilon 4$  alone ( $p < 0.05$ ; [Figure 2D](#)). All NN classification results are summarized in [Table 2](#). NN performance scores of using from 1 to 100 top proteins with or without age and *APOE* status to classify each outcome are shown in [Supplementary Figure 1](#).

TABLE 2 Summary of AUC scores (95% CI) of NN classification.

	Number of subjects (high/low)	Age + <i>APOE</i> only	Proteins only	Proteins + age + <i>APOE</i>
A $\beta$	503/378	0.748 (0.745–0.750)	0.727 (0.722–0.732)	0.782 (0.779–0.785)
P-tau	373/414	0.662 (0.658–0.666)	0.626 (0.615–0.636)	0.674 (0.669–0.678)
T-tau	415/376	0.710 (0.708–0.713)	0.668 (0.660–0.675)	0.734 (0.729–0.738)
AT(N)	298/229	0.795 (0.793–0.798)	0.770 (0.765–0.774)	0.831 (0.827–0.835)

AUC, area under the receiver operating characteristic curve; CI, confidence interval; NN, neural networks; *APOE*, apolipoprotein E; A $\beta$ , amyloid- $\beta$ ; p-tau, phosphorylated-tau; t-tau, total tau; AT(N), amyloid/tau/neurodegeneration.

## Enriched terms

Aggregately, 35 proteins were nominated for the pathway enrichment analysis. They were matched to 34 unique genes. The background list consisted of 3,306 unique proteins in UniProt identifiers, which were matched to 3,268 genes. A total of 11 pathways were identified ([Table 3](#)). They were grouped into five clusters based on their membership similarities. Each cluster was represented by its most significant pathway: human immunodeficiency virus 1 infection (uncorrected  $p < 0.001$ ), p53 signaling pathway (uncorrected  $p = 0.001$ ), phosphoinositide-3-kinase–protein kinase B/Akt (PI3K–Akt) signaling pathway (uncorrected  $p = 0.009$ ), complement and coagulation cascades (uncorrected  $p = 0.018$ ), and mitogen-activated protein kinase (MAPK) signaling pathway (uncorrected  $p = 0.045$ ). The enriched pathways were not significant after correction for multiple comparisons (corrected  $p > 0.05$ ). The protein–protein interaction analysis revealed three significant terms ([Table 4](#)): PI3K–Akt signaling pathway ( $p < 0.001$ ), complement and coagulation cascades ( $p < 0.001$ ), and proteoglycans in cancer ( $p < 0.005$ ).

## Discussion

In this study, we used regression analysis and NNs to obtain the optimal sets of protein features that discriminated high and low AT(N) burdens in AD. Age and *APOE*  $\epsilon 4$  status alone achieved an AUC of 0.748 in predicting A $\beta$ , an AUC of 0.662 in predicting p-tau, an AUC of 0.710 in predicting t-tau, and an AUC of 0.795 in predicting AT(N) abnormality. The addition of proteins significantly improved prediction of A $\beta$  (AUC = 0.782), p-tau (AUC = 0.674), t-tau (AUC = 0.734) and AT(N) profiles (AUC = 0.831).

We selected the variables age and *APOE* genotype for comparison as advanced age and presence of the *APOE*  $\epsilon 4$  allele are two of the strongest risk factors for AD ([Riedel et al., 2016](#)). In addition, our previous study confirmed that among age, sex, education, and *APOE* genotype, the combination of these two characteristics best predicts A $\beta$  pathologies in the current cohort ([Shi Y. et al., 2019](#)). The previous study also classified amyloid burden using plasma proteomics and obtained similar results: the combination of age, *APOE*  $\epsilon 4$  status, and proteins achieved the highest AUC score of 0.78, outperforming demographic variables alone. But as different protein ranking methods were employed, our panel of 11 proteins predicting A $\beta$  has little overlap with the previously identified 44 proteins and provides novel candidates for validation. In the previous study, top proteins were selected through Lasso which uses a regularization technique to choose features that correlate with the outcome but not with each other, thus reducing redundant inputs. While this method can enhance prediction accuracy in an independent testing set, it may exclude proteins that have linearly redundant association with the outcome but nonlinear effects that can be detected by neural networks. As logistic regression does not penalize feature collinearity, the

**TABLE 3** Pathway enrichment analysis revealed 11 significantly enriched pathways.

Term ID	Description	P-value (uncorrected)	P-value (corrected)
hsa05170	Human immunodeficiency virus 1 infection	< 0.001	0.203
hsa04218	Cellular senescence	0.009	0.641
hsa05145	Toxoplasmosis	0.015	0.754
hsa05014	Amyotrophic lateral sclerosis	0.036	1.000
hsa04115	p53 signaling pathway	0.001	0.209
hsa04114	Oocyte meiosis	0.005	0.524
hsa04151	PI3K-Akt signaling pathway	0.009	0.641
hsa05202	Transcriptional misregulation in cancer	0.013	0.747
hsa04610	Complement and coagulation cascades	0.018	0.770
hsa04936	Alcoholic liver disease	0.029	1.000
hsa04010	MAPK signaling pathway	0.045	1.000

They are grouped into five clusters based on membership similarities. PI3K-Akt, phosphoinositide-3-kinase-protein kinase B/Akt; MAPK, mitogen-activated protein kinase.

**TABLE 4** Protein-protein interaction enrichment analysis revealed three significantly enriched terms.

Term ID	Description	P-value
hsa04151	PI3K-Akt signaling pathway	<0.001
hsa04610	Complement and coagulation cascades	<0.001
hsa05205	Proteoglycans in cancer	<0.005

PI3K-Akt, phosphoinositide-3-kinase-protein kinase B/Akt.

proteins selected using this method may retain meaningful information for neural network training. Their performance corroborates the utility of plasma proteomics in predicting A $\beta$  pathology demonstrated in other recent studies (Ashton et al., 2019; Park et al., 2019; Westwood et al., 2020). Considering that AD is a complex disorder with mixed pathologies, we further evaluated the potential of plasma proteomics in predicting the other two components of the AT(N) framework. In accordance with previous findings (Shi et al., 2021), our results suggest that blood-based protein panels can reflect brain tau burden and neurodegeneration in addition to A $\beta$  abnormality. Finally, a panel of proteins showed satisfactory performance in predicting combined AT(N) profiles, supporting the potential of plasma proteomics to act as comprehensive biomarkers for core AD pathologies. Interestingly, the statistical significance measures of the association between proteins and each AD-related outcomes have high correlations, suggesting that these pathological features are closely related, which is consistent with our previous finding showing that A $\beta$  has a causal relationship with tau pathology (Shi L. et al., 2019).

We performed a pathway enrichment analysis on proteins in identified panels and found 5 clusters of pathways. One of them is the complement and coagulation cascades, which have been

reported previously in a systematic review of blood-based protein biomarkers (Kiddle et al., 2014). Neuroinflammation is implicated as having substantial involvement in the pathogenesis of AD (Heneka et al., 2015). As part of the innate immune system, the complement system is activated by A $\beta$  deposits and in turn damages the neurons *via* self-attack (McGeer and McGeer, 2002). Its activation is accompanied by an upregulation of coagulation factors (Amara et al., 2010) associated with the neurovascular damages observed in AD brain. A previous study supports the association between the peripheral activation of this pathway and AD-specific pathologies (Pillai et al., 2019). Notably, the complement protein C4 was nominated in multiple panels. A previous study has found that C4 could discriminate rapidly and slowly progressing AD (Thambisetty et al., 2010), suggesting that it might be indicative of AD severity and is a potentially promising biomarker for early stages of AD. Another significant pathway identified is the MAPK pathway, which has been recognized in the AlzPathway, a comprehensive map of pathways related to AD (Mizuno et al., 2012). In AD, activation of the serine/threonine MAPK in response to extracellular stimuli promotes neuronal apoptosis (Munoz and Ammit, 2010). C-Jun N-terminal kinases (JNK) and p38, two members of the MAPK family, are involved in the hyperphosphorylation of tau (Goedert et al., 1997; Reynolds et al., 2000). JNK is also thought to regulate the phosphorylation and degradation of amyloid precursor proteins (Muresan and Muresan, 2007; Colombo et al., 2009). Our results provide further evidence for the dysregulation of the MAPK cascade proteins in AD.

This study leveraged the advancement of ML. While traditional ML models have been predominant in previous investigation of AD biomarkers, deep NNs are expected to play a more significant role moving forward. They are capable of detecting complex nonlinear patterns in raw data and are highly sensitive to the relevance of information received (LeCun et al., 2015). Recent studies have demonstrated the superior performance of deep learning in detecting AD disease stage and predicting longitudinal progression of AD using multimodal information (Venugopalan et al., 2021; El-Sappagh et al., 2022). In agreement, our study indicates the great potential of deep learning approaches to capture the complexity of blood-based omics data and facilitate the discovery of candidate biomarkers. While the performance of NNs in this study did not exceed that of the support vector machine used in our previous study (Shi L. et al., 2019), this might result from scarcity of data as training of deep learning models typically benefits from extremely large sample sizes (Ellis and Morgan, 1999). A similar study found that the more conventional random forests outperformed deep learning models in differentiating AD from CN using plasma metabolomics and reached the same conclusion (Stamate et al., 2019). While NN behavior has been much less explored in bioinformatic data, extensive deep learning research in imaging, video, audio, and natural language processing has consistently shown that model performance increases with data size, a phenomenon now known as the scaling laws (Brown et al., 2020; Wei et al., 2022). In light of these observations, it is plausible that a similar effect would

become apparent in omics as data sizes grow orders of magnitude larger than thousands of samples still typically used in present-day studies. Model architecture should also be considered in the future development of AD biomarkers. More complex NNs can be adopted to take advantage of an increasing quantity of heterogeneous clinical and biological data. Recently, TabNet has been proposed as a NN architecture specializing in tabular data processing by applying sequential attention to select the best features at each decision step (Arik and Pfister, 2021). It would be interesting to see whether innovative models like TabNet could be applied to explore the predictive power of putative biomarkers and integrated into the biomarker development pipeline.

One limitation of the current study is that no data was withheld for validation of the effect of the nominated proteins. However, this helps minimize overfitting and optimize the generalizability of those proteins at the discovery stage. To assess their applicability in the larger population, future validation using independent cohorts is preferred. In addition, it is important to note that participants with AD were significantly older than MCI or CN participants in this study, so age and age-related changes in plasma profiles might contribute to most of the predictive accuracy. As development of AD is not always associated with age in the larger population, it is important to test the validity of candidate biomarkers among patients and controls of similar ages.

In conclusion, the opportunity of the clinical implementation of blood-based biomarkers for AD is exciting. The current study supports the use of proteomics measured by SOMAscan for the discovery blood-based biomarkers. In addition, NNs show great utility in predicting disease pathologies from proteomics which encourages the adoption of more advanced ML approaches in future investigation. Using these state-of-the-art technologies, we identified several proteins that are involved in AD-related pathways and can potentially serve as prescreening tools for the early detection of AD-specific pathologies when combined with demographic information.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the dataset analyzed for this study is available upon request *via* the EMIF-AD Catalog (<https://emif-catalogue.eu>) after approval of the research question by all parent cohorts and the EMIF-AD team. Requests to access these datasets should be directed to EMIF-AD, <https://emif-catalogue.eu>.

## Ethics statement

The studies involving human participants were reviewed and approved by Aristotle University of Thessaloniki Medical School Ethics Committee; Ethics Committee of the Medical Faculty Mannheim, University of Heidelberg; Ethic and Clinical Research Committee Donostia; Ethics Committee Inserm and Aix Marseille University; The Healthcare Ethics Committee of the Hospital Clinic;

Central Clinical Research and Clinical Trials Unit (UICEC Sant Pau); INSERM Ethical Committee; Ethic Committee of the IRCCS San Giovanni di Dio FBF; Comitato Etico IRCCS Pascale - Napoli; Ethics Committee at Karolinska Institutet; Ethische commissie onderzoek UZ/KU Leuven; Research Ethics Committee Lausanne University Hospital; Medical Ethical Committee Maastricht University Medical Center; Committee on Health Research Ethics, Region of Denmark; Ethics Committee of Mediterranean University; University of Lille Ethics Committee; Ethical Committee at the Medical Faculty, Leipzig University; Ethical Committee at the Medical Faculty, University Hospital Essen; Ethics Committee University of Antwerp; Ethical Committee of University of Genoa; Ethics Committee, University of Gothenburg; Human Ethics Committee of the University of Perugia; and Medical Ethics Committee VU Medical Center. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

AN-H, LS, and YZ contributed to conception and design of the study. YZ analyzed the data, interpreted the results, and drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was conducted as part of the EMIF-AD project which has received support from the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement no. 115372, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in-kind contribution. The authors declare that they have received funding from Astra Zeneca (SL) and Janssen (SL and ANH). The funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication. The DESCRIPA study was funded by the European Commission within the 5th framework program (QLRT-2001-2455). The EDAR study was funded by the European Commission within the 5th framework program (contract # 37670). The San Sebastian GAP study was partially funded by the Department of Health of the Basque Government (allocation 17.0.1.08.12.0000.2.454.01.41142.001.H). The research at VIB-CMN was funded in part by the University of Antwerp Research Fund. LS is funded by the Virtual Brain Cloud from European commission (grant no. H2020-SC1-DTH-2018-1). HZ is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018–02532), the European Research Council (#681712 and #101053962), Swedish State Support for Clinical Research (#ALFGBG-71320), the Alzheimer Drug Discovery Foundation (ADDF), United States (#201809–2016862), the AD Strategic Fund and the Alzheimer's Association (#ADSF-21-831376-C, #ADSF-21-831381-C, and #ADSF-21-831377-C), the Bluefield Project, the Olav Thon Foundation, the Erling-Persson



Family Foundation, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2022-0270), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no 860197 (MIRIADE), the European Union Joint Programme – Neurodegenerative Disease Research (JPND2021-00694), the UK Dementia Research Institute at UCL (UKDRI-1003), and the Lausanne cohort was supported by grants from the Swiss National Science Foundation (SNF 320030\_141179), Synopsis Foundation – Dementia Research Switzerland (grant no. 2017-PI01). This work was supported by the Centre for Artificial Intelligence in Precision Medicines of the University of Oxford and King Abdulaziz University.

## Conflict of interest

SL is named as an inventor on biomarker intellectual property protected by Proteome Sciences and Kings College London unrelated to the current study and within the past 5 years has advised for Optum labs, Merck, SomaLogic and been the recipient of funding from AstraZeneca and other companies *via* the IMI funding scheme. SL is employed by company Janssen. HZ has served at scientific advisory boards and/or as a consultant for Abbvie, Acumen, Alector, ALZPath, Annexon, Apellis, Artery Therapeutics, AZTherapies, CogRx, Denali, Eisai, Nervgen, Novo Nordisk, Passage Bio, Pinteon Therapeutics, Red Abbey Labs, reMYND, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics, and Wave, has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, Biogen, and Roche, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS),

which is a part of the GU Ventures Incubator Program, all unrelated to this study. AL has served at scientific advisory boards of Biogen, Fujirebio Europe, Eli Lilly, Grifols, Novartis, Nutricia, Roche, and Otsuka and is the inventor of a patent on synaptic markers in CSF, all unrelated to this study. JP has served at scientific advisory boards of Fujirebio Europe, Eli Lilly, and Nestlé Institute of Health Sciences, all unrelated to this study. SE has received unrestricted research grants from Janssen Pharmaceutica and ADx Neurosciences and has served at scientific advisory boards of Biogen, Eisai, Novartis, Nutricia/Danone, and Roche, all unrelated to this study.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2022.1040001/full#supplementary-material>

## References

- Amara, U., Flierl, M. A., Rittirsch, D., Klos, A., Chen, H., Acker, B., et al. (2010). Molecular intercommunication between the complement and coagulation systems. *J. Immunol.* 185, 5628–5636. doi: 10.4049/jimmunol.0903678
- Apostolova, L. G. (2016). Alzheimer disease. *Continuum* 22, 419–434. doi: 10.1212/CON.0000000000000307
- Arik, S. O., and Pfister, T. (2021). Tabnet: Attentive interpretable tabular learning. arXiv [Preprint]. 35, 6679–6687.
- Ashton, N. J., Nevado-Holgado, A. J., Barber, I. S., Lynham, S., Gupta, V., Chatterjee, P., et al. (2019). A plasma protein classifier for predicting amyloid burden for preclinical Alzheimer's disease. *Science. Advances* 5:eaau7220. doi: 10.1126/sciadv.aau7220
- Ashton, N. J., Pascoal, T. A., Karikari, T. K., Benedet, A. L., Lantero-Rodriguez, J., Brinkmalm, G., et al. (2021). Plasma p-tau231: a new biomarker for incipient Alzheimer's disease pathology. *Acta Neuropathol.* 141, 709–724. doi: 10.1007/s00401-021-02275-6
- Beltrán, J. F., Wahba, B. M., Hose, N., Shasha, D., and Kline, R. P. For the Alzheimer's Disease Neuroimaging Initiative (2020). Inexpensive, non-invasive biomarkers predict Alzheimer transition using machine learning analysis of the Alzheimer's disease neuroimaging (ADNI) database. *PLoS One* 15:e0235663. doi: 10.1371/journal.pone.0235663
- Bos, I., Vos, S., Vandenbergh, R., Scheltens, P., Engelborghs, S., Frisoni, G., et al. (2018). The EMIF-AD multimodal biomarker discovery study: design, methods and cohort characteristics. *Alzheimers Res. Ther.* 10:64. doi: 10.1186/s13195-018-0396-5
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* 33, 1877–1901. doi: 10.48550/arXiv.2005.14165
- Colombo, A., Bastone, A., Ploia, C., Scip, A., Salmona, M., Forloni, G., et al. (2009). JNK regulates APP cleavage and degradation in a model of Alzheimer's disease. *Neurobiol. Dis.* 33, 518–525. doi: 10.1016/j.nbd.2008.12.014
- Curtis, C., Gamez, J. E., Singh, U., Sadowsky, C. H., Villena, T., Sabbagh, M. N., et al. (2015). Phase 3 trial of flutemetamol labeled with radioactive fluorine 18 imaging and neuritic plaque density. *JAMA Neurol.* 72, 287–294. doi: 10.1001/jamaneurol.2014.4144
- de Almeida, S. M., Shumaker, S. D., LeBlanc, S. K., Delaney, P., Marquie-Beck, J., Ueland, S., et al. (2011). Incidence of post-dural puncture headache in research volunteers. *Headache* 51, 1503–1510. doi: 10.1111/j.1526-4610.2011.01959.x
- Dey, K. K., Wang, H., Niu, M., Bai, B., Wang, X., Li, Y., et al. (2019). Deep undepleted human serum proteome profiling toward biomarker discovery for Alzheimer's disease. *Clin. Proteomics* 16:16. doi: 10.1186/s12014-019-9237-1
- Durstewitz, D., Koppe, G., and Meyer-Lindenberg, A. (2019). Deep neural networks in psychiatry. *Mol. Psychiatry* 24, 1583–1598. doi: 10.1038/s41380-019-0365-9
- Ellis, D., and Morgan, N. (1999). Size matters: an empirical study of neural network training for large vocabulary continuous speech recognition. *IEEE*, 1013–1016. doi: 10.1109/ICASSP.1999.759875
- El-Sappagh, S., Saleh, H., Ali, F., Amer, E., and Abuhmed, T. (2022). Two-stage deep learning model for Alzheimer's disease detection and prediction of the mild cognitive impairment time. *Neural Comput. Applic.* 34, 14487–14509. doi: 10.1007/s00521-022-07263-9
- Goedert, M., Hasegawa, M., Jakes, R., Lawler, S., Cuenda, A., and Cohen, P. (1997). Phosphorylation of microtubule-associated protein tau by stress-activated protein kinases. *FEBS Lett.* 409, 57–62. doi: 10.1016/S0014-5793(97)00483-3

- Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E., et al. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. *Nat. Prec.* 5, e15004–e15004. doi: 10.1038/npre.2010.4538.1
- Goudey, B., Fung, B. J., Schieber, C., Faux, N. G., Weiner, M. W., Aisen, P., et al. (2019). A blood-based signature of cerebrospinal fluid A $\beta$ 1–42 status. *Sci. Rep.* 9:4163. doi: 10.1038/s41598-018-37149-7
- Heneka, M. T., Carson, M. J., Khoury, J. E., Landreth, G. E., Brosseron, F., Feinstein, D. L., et al. (2015). Neuroinflammation in Alzheimer's disease. *Lancet Neurol.* 14, 388–405. doi: 10.1016/S1474-4422(15)70016-5
- Hye, A., Lynham, S., Thambisetty, M., Causevic, M., Campbell, J., Byers, H. L., et al. (2006). Proteome-based plasma biomarkers for Alzheimer's disease. *Brain* 129, 3042–3050. doi: 10.1093/brain/awl279
- Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., et al. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol.* 12, 207–216. doi: 10.1016/S1474-4422(12)70291-0
- Jiang, Y., Zhou, X., Ip, F. C., Chan, P., Chen, Y., Lai, N. C., et al. (2022). Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. *Alzheimers Dement.* 18, 88–102. doi: 10.1002/alz.12369
- Karaglani, M., Gourlia, K., Tsamardinos, I., and Chatzaki, E. (2020). Accurate blood-based diagnostic biosignatures for Alzheimer's disease via automated machine learning. *J. Clin. Med.* 9:1036. doi: 10.3390/jcm9093016
- Kiddle, S. J., Sattler, M., Proitsi, P., Simmons, A., Westman, E., Bazenot, C., et al. (2014). Candidate blood proteome markers of Alzheimer's disease onset and progression: a systematic review and replication study. *J. Alzheimers Dis.* 38, 515–531. doi: 10.3233/JAD-130380
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Li, Z., Jiang, X., Wang, Y., and Kim, Y. (2021). Applied machine learning in Alzheimer's disease research: omics, imaging, and clinical data. *Emerg. Topics Life Sci.* 5, 765–777. doi: 10.1042/ETLS20210249
- Lin, C.-H., Chiu, S.-I., Chen, T.-F., Jang, J.-S. R., and Chiu, M.-J. (2020). Classifications of neurodegenerative disorders using a multiplex blood biomarkers-based machine learning model. *Int. J. Mol. Sci.* 21:6914. doi: 10.3390/ijms21186914
- Lopez, M. F., Mikulskis, A., Kuzdzal, S., Bennett, D. A., Kelly, J., Golenko, E., et al. (2005). High-resolution serum proteomic profiling of Alzheimer disease samples reveals disease-specific, carrier-protein-bound mass signatures. *Clin. Chem.* 51, 1946–1954. doi: 10.1373/clinchem.2005.053090
- McGeer, P. L., and McGeer, E. G. (2002). The possible role of complement activation in Alzheimer disease. *Trends Mol. Med.* 8, 519–523. doi: 10.1016/S1471-4914(02)02422-X
- Milà-Alomà, M., Ashton, N. J., Shekari, M., Salvadó, G., Ortiz-Romero, P., Montoliu-Gaya, L., et al. (2022). Plasma p-tau231 and p-tau217 as state markers of amyloid- $\beta$  pathology in preclinical Alzheimer's disease. *Nat. Med.* 28, 1797–1801. doi: 10.1038/s41591-022-01925-w
- Mizuno, S., Iijima, R., Ogishima, S., Kikuchi, M., Matsuoka, Y., Ghosh, S., et al. (2012). AlzPathway: a comprehensive map of signaling pathways of Alzheimer's disease. *BMC Syst. Biol.* 6:52. doi: 10.1186/1752-0509-6-52
- Munoz, L., and Ammit, A. J. (2010). Targeting p38 MAPK pathway for the treatment of Alzheimer's disease. *Neuropharmacology* 58, 561–568. doi: 10.1016/j.neuropharm.2009.11.010
- Muresan, Z., and Muresan, V. (2007). The amyloid- $\beta$  precursor protein is phosphorylated via distinct pathways during differentiation, mitosis, stress, and degeneration. *Mol. Biol. Cell* 18, 3835–3844. doi: 10.1091/mbc.e06-07-0625
- O'Bryant, S. E., Edwards, M., Johnson, L., Hall, J., Villarreal, A. E., Britton, G. B., et al. (2016). A blood screening test for Alzheimer's disease. *Alzheimers Dement.* 3, 83–90. doi: 10.1016/j.dadm.2016.06.004
- Palmqvist, S., Zetterberg, H., Mattsson, N., Johansson, P., For the Alzheimer's Disease Neuroimaging Initiative, Minthon, L., et al. (2015). Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. *Neurology* 85, 1240–1249. doi: 10.1212/WNL.0000000000001991
- Park, J.-C., Han, S.-H., Lee, H., Jeong, H., Byun, M. S., Bae, J., et al. (2019). Prognostic plasma protein panel for A $\beta$  deposition in the brain in Alzheimer's disease. *Prog. Neurobiol.* 183:101690. doi: 10.1016/j.pneurobio.2019.101690
- Pérez-Grijalva, V., Arbizu, J., Romero, J., Prieto, E., Pesini, P., Sarasa, L., et al. (2019). Plasma A $\beta$ 42/40 ratio alone or combined with FDG-PET can accurately predict amyloid-PET positivity: a cross-sectional analysis from the AB255 study. *Alzheimers Res. Ther.* 11:96. doi: 10.1186/s13195-019-0549-1
- Pillai, J. A., Maxwell, S., Bena, J., Bekris, L. M., Rao, S. M., Chance, M., et al. (2019). Key inflammatory pathway activations in the MCI stage of Alzheimer's disease. *Ann. Clin. Transl. Neurol.* 6, 1248–1262. doi: 10.1002/acn3.50827
- Ray, S., Britschgi, M., Herbert, C., Takeda-Uchimura, Y., Boxer, A., Blennow, K., et al. (2007). Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat. Med.* 13, 1359–1362. doi: 10.1038/nm1653
- Reynolds, C. H., Betts, J. C., Blackstock, W. P., Nebreda, A. R., and Anderton, B. H. (2000). Phosphorylation sites on tau identified by nano-electrospray mass spectrometry: differences in vitro between the mitogen-activated protein kinases ERK2, c-Jun N-terminal kinase and P38, and glycogen synthase kinase-3 $\beta$ . *J. Neurochem.* 74, 1587–1595. doi: 10.1046/j.1471-4159.2000.0741587.x
- Riedel, B. C., Thompson, P. M., and Brinton, R. D. (2016). Age, APOE and sex: triad of risk of Alzheimer's disease. *J. Steroid Biochem. Mol. Biol.* 160, 134–147. doi: 10.1016/j.jsmb.2016.03.012
- Sattler, M., Kiddle, S. J., Newhouse, S., Proitsi, P., Nelson, S., Williams, S., et al. (2014). Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimers Dement.* 10, 724–734. doi: 10.1016/j.jalz.2013.09.016
- Shi, Y., Lu, X., Zhang, L., Shu, H., Gu, L., Wang, Z., et al. (2019). Potential value of plasma amyloid- $\beta$ , total tau, and neurofilament light for identification of early Alzheimer's disease. *ACS Chem. Neurosci.* 10, 3479–3485. doi: 10.1021/acscchemneuro.9b00095
- Shi, L., Westwood, S., Baird, A. L., Winchester, L., Dobricic, V., Kilpert, F., et al. (2019). Discovery and validation of plasma proteomic biomarkers relating to brain amyloid burden by SOMAscan assay. *Alzheimers Dement.* 15, 1478–1488. doi: 10.1016/j.jalz.2019.06.4951
- Shi, L., Winchester, L. M., Westwood, S., Baird, A. L., Anand, S. N., Buckley, N. J., et al. (2021). Replication study of plasma proteins relating to Alzheimer's pathology. *Alzheimers Dement.* 17, 1452–1464. doi: 10.1002/alz.12322
- Stamate, D., Kim, M., Proitsi, P., Westwood, S., Baird, A., Nevado-Holgado, A., et al. (2019). A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: results from the European medical information framework for Alzheimer disease biomarker discovery cohort. *Alzheimers Dement.* 5, 933–938. doi: 10.1016/j.trci.2019.11.001
- Strozyk, D., Blennow, K., White, L. R., and Launer, L. J. (2003). CSF A $\beta$  42 levels correlate with amyloid-neuropathology in a population-based autopsy study. *Neurology* 60, 652–656. doi: 10.1212/01.WNL.0000046581.81650.D0
- Thambisetty, M., Simmons, A., Velayudhan, L., Hye, A., Campbell, J., Zhang, Y., et al. (2010). Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. *Arch. Gen. Psychiatry* 67, 739–748. doi: 10.1001/archgenpsychiatry.2010.78
- Venugopalan, J., Tong, L., Hassanzadeh, H. R., and Wang, M. D. (2021). Multimodal deep learning models for early detection of Alzheimer's disease stage. *Sci. Rep.* 11:3254. doi: 10.1038/s41598-020-74399-w
- Voyle, N., Baker, D., Burnham, S. C., Covin, A., Zhang, Z., Sangurdekar, D. P., et al. (2015). Blood protein markers of neocortical amyloid- $\beta$  burden: a candidate study using SOMAscan technology. *J. Alzheimers Dis.* 46, 947–961. doi: 10.3233/JAD-150020
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., et al. (2022). Emergent abilities of large language models. arXiv: 2206.07682 [Preprint].
- Westwood, S., Baird, A. L., Anand, S. N., Nevado-Holgado, A. J., Kormilitzin, A., Shi, L., et al. (2020). Validation of plasma proteomic biomarkers relating to brain amyloid burden in the EMIF-Alzheimer's disease multimodal biomarker discovery cohort. *J. Alzheimers Dis.* 74, 213–225. doi: 10.3233/JAD-190434
- Whelan, C. D., Mattsson, N., Nagle, M. W., Vijayaraghavan, S., Hyde, C., Janelidze, S., et al. (2019). Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's disease. *Acta Neuropathol. Commun.* 7:169. doi: 10.1186/s40478-019-0795-2
- Zhao, X., Kang, J., Svetnik, V., Warden, D., Wilcock, G., David Smith, A., et al. (2020). A machine learning approach to identify a circulating microRNA signature for Alzheimer disease. *J. Appl. Lab. Med.* 5, 15–28. doi: 10.1373/jalm.2019.029595
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascap provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10, 1–10. doi: 10.1038/s41467-019-09234-6