Short Report

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

Daniel Stamate[a,b,c], Min Kim[d], Petroula Proitsi[e], Sarah Westwood[f], Alison Baird[f], Alejo Nevado-Holgado[f], Abdul Hye[e], Isabelle Bos[g,h], Stephanie J. B. Vos[g], Rik Vandenberghe[h], Charlotte E. Teunissen[i], Mara Ten Kate[h,i], Philip Scheltens[h], Silvy Gabel[j,k,l], Karen Meersmans[k,l], Olivier Blin[m], Jill Richardson[n], Ellen De Roeck[o,p,q], Sebastiaan Engelborghs[p,q,r], Kristel Sleegers[q,s], Régis Bordet[t], Lorena Ramit[u], Petronella Kettunen[v], Magda Tsolaki[w], Frans Verhey[g], Daniel Alcolea[x], Alberto Lléo[x], Gwendoline Peyratout[y], Mikel Tainta[z], Peter Johannsen[aa], Yvonne Freund-Levi[e,bb], Lutz Frölich[cc], Valerija Dobricic[dd], Giovanni B. Frisoni[ee,ff], José L. Molinuevo[t,gg], Anders Wallin[hh], Julius Popp[y,ii], Pablo Martinez-Lage[z], Lars Bertram[dd,jj], Kaj Blennow[kk,ll], Henrik Zetterberg[kk,ll,mm,nn], Johannes Streffer[oo], Pieter J. Visser[g,h], Simon Lovestone[f,pp], Cristina Legido-Quigley[d,qq,*]

[a]Division of Population Health, Health Services Research and Primary Care, University of Manchester, Manchester, UK
[b]Data Science & Soft Computing Lab, London, UK
[c]Computing Department, Goldsmiths College, University of London, London, UK
[d]Steno Diabetes Center Copenhagen, Gentofte, Denmark
[e]Institute of Psychiatry, Psychology and Neuroscience, Maurice Wohl Clinical Neuroscience Institute, King's College London, London, UK
[f]Department of Psychiatry, University of Oxford, Oxford, UK
[g]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Alzheimer Centrum Limburg, Maastricht University, Maastricht, the Netherlands
[h]Department of Neurology, Alzheimer Center Amsterdam, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, the Netherlands
[i]Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, the Netherlands
[j]Department of Clinical Chemistry, Neurochemistry Laboratory, Amsterdam Neuroscience, Amsterdam University Medical Centers, Vrije Universiteit, the Netherlands
[k]University Hospital Leuven, Leuven, Belgium
[l]Department of Neurosciences, Laboratory for Cognitive Neurology, KU Leuven, Belgium
[m]AIX Marseille University, INS, Ap-hm, Marseille, France
[n]Neurosciences Therapeutic Area, GlaxoSmithKline R&D, Stevenage, UK
[o]Faculty of Psychology & Educational Sciences Vrije Universiteit Brussel (VUB), Brussels, Belgium
[p]Reference Center for Biological Markers of Dementia (BIODEM), University of Antwerp, Antwerp, Belgium
[q]Institute Born-Bunge, University of Antwerp, Antwerp, Belgium
[r]Department of Neurology, UZ Brussel and Center for Neurosciences, Vrije Universiteit Brussel (VUB), Brussels, Belgium
[s]Neurodegenerative Brain Diseases Group, Center for Molecular Neurology, VIB, Belgium
[t]University of Lille, Inserm, CHU Lille, Lille, France
[u]Alzheimer's Disease & Other Cognitive Disorders Unit, Hospital Clínic-IDIBAPS, Barcelona, Spain
[v]Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden
[w]1st Department of Neurology, AHEPA University Hospital, Makedonia, Thessaloniki, Greece

*Corresponding author. Tel.: +4530913083.
E-mail address: cristina.legido.quigley@regionh.dk

[x]*Memory Unit, Neurology Department, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain*
[y]*University Hospital of Lausanne, Lausanne, Switzerland*
[z]*Center for Research and Advanced Therapies, Fundacion CITA-alzheimer Fundazioa, Donostia/San Sebastian, Spain*
[aa]*Danish Dementia Research Centre, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark*
[bb]*Department of Neurobiology, Caring Sciences and Society (NVS), Division of Clinical Geriatrics, Karolinska Institute, and Department of Geriatric Medicine,*
Karolinska University Hospital Huddinge, Stockholm, Sweden
[cc]*Department of Geriatric Psychiatry, Zentralinstitut für Seelische Gesundheit, University of Heidelberg, Mannheim, Germany*
[dd]*Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and Cardiogenetics, University of Lübeck, Lübeck, Germany*
[ee]*University of Geneva, Geneva, Switzerland*
[ff]*IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy*
[gg]*Barcelona Beta Brain Research Center, Unversitat Pompeu Fabra, Barcelona, Spain*
[hh]*Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden*
[ii]*Department of Mental Health and Psychiatry, Geriatric Psychiatry, Geneva University Hospitals, Geneva, Switzerland*
[jj]*Department of Psychology, University of Oslo, Oslo, Norway*
[kk]*Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, University of Gothenburg, Mölndal, Sweden*
[ll]*Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden*
[mm]*UK Dementia Research Institute at UCL, London, UK*
[nn]*Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK*
[oo]*Reference Center for Biological Markers of Dementia (BIODEM), Institute Born-Bunge, University of Antwerp, Antwerp, Belgium*
[pp]*Janssen-Cilag UK Ltd, Oxford, UK*
[qq]*Institute of Pharmaceutical Science, King's College London, London, UK*

Abstract

Introduction: Machine learning (ML) may harbor the potential to capture the metabolic complexity in Alzheimer Disease (AD). Here we set out to test the performance of metabolites in blood to categorize AD when compared to CSF biomarkers.

Methods: This study analyzed samples from 242 cognitively normal (CN) people and 115 with AD-type dementia utilizing plasma metabolites (n = 883). Deep Learning (DL), Extreme Gradient Boosting (XGBoost) and Random Forest (RF) were used to differentiate AD from CN. These models were internally validated using Nested Cross Validation (NCV).

Results: On the test data, DL produced the AUC of 0.85 (0.80–0.89), XGBoost produced 0.88 (0.86–0.89) and RF produced 0.85 (0.83–0.87). By comparison, CSF measures of amyloid, p-tau and t-tau (together with age and gender) produced with XGBoost the AUC values of 0.78, 0.83 and 0.87, respectively.

Discussion: This study showed that plasma metabolites have the potential to match the AUC of well-established AD CSF biomarkers in a relatively small cohort. Further studies in independent cohorts are needed to validate whether this specific panel of blood metabolites can separate AD from controls, and how specific it is for AD as compared with other neurodegenerative disorders.

1. Introduction

At present, the diagnosis of Alzheimer disease–type dementia (AD) is based on protein biomarkers in cerebrospinal fluid (CSF) and brain imaging together with a battery of cognition tests. Diagnostic tools based on CSF collection are invasive while brain-imaging tools are still costly, and therefore, there is a need to identify noninvasive tools for early detection as well as for measuring disease progression.

In recent years, an increasing number of studies have examined blood metabolites as potential AD biomarkers [1–4]. The advantages of looking at blood metabolites are that they are easily accessible but also that they represent an essential aspect of the phenotype of an organism and hence might act as a molecular fingerprint of disease

progression [5,6]. Therefore, blood AD markers could potentially aid early diagnosis and recruitment for trials.

Here we utilized data generated as part of the European Medical Information Framework for AD Multimodal Biomarker Discovery (EMIF-AD) previously reported in full in Kim et al. [7]. As discussed in that paper, metabolite levels were measured using liquid chromatography–mass spectroscopy (LC-MS) to cover ca. Eight hundred metabolites and these metabolites related to CSF biomarkers of AD commonly used in clinical research including trials, and increasingly in clinical practice, as part of the diagnostic work up. Here we explore the potential of different Machine Learning (ML) algorithms to identify those individuals with AD from dataset and to compare the effectiveness of blood-based metabolites as an indicator of clinical diagnosis to that of CSF markers. In this study we employed two state-of-the-

art ML algorithms—Deep Learning (DL) and Extreme Gradient Boosting (XGBoost)—and compared these to the more commonly utilized Random Forest (RF) algorithm.

2. Methods

This study accessed data previously generated from 242 samples from cognitively normal (CN) individuals and 115 from people with AD-type dementia (AD) samples in which diagnosis was based on clinical diagnosis. Details on the subjects, clinical and cognitive data, as well as measurements of AD pathological markers have been described elsewhere [7,8]. The metabolomics data employed here was accessed in the EMIF-AD portal and the acquisition and processing details can be found via open access in [7]. In short, the EMIF-AD cohort is a collated cohort making use of existing data and samples collected in 11 different studies across Europe, with the aim to discover novel diagnostic and prognostic markers for pre-dementia AD.

In the current study, the main objective was to use state-of-the-art ML classification algorithms to build CN versus. AD predictive models using blood metabolites. For this purpose, we employed DL and XGBoost. Additionally we also employed the more popularly used RF algorithm. These models were compared in terms of binary classifiers with Area Under the Curve (AUC) in Receiver Operating Characteristic (ROC) curves.

The metabolites with more than 45% missing values were discarded. The remaining missing values were handled with imputation methods based on the k-nearest neighbor (RF and DL), or internally by the classification algorithm (XGBoost). Models were built and evaluated using a Nested Cross Validation (NCV), which used 9/10 data folds for model training and optimization in an inner cross-validation, and 1/10 data folds for model testing in an outer cross-validation. The process was repeated 10 times, for each of the test data folds.

The analysis was further extended by assessing the stability of the AUC performance with Monte Carlo (MC) simulations consisting of 50 repeated similar NCV experiments. As such, multiple models were built on multiple samples in the NCV and MC, using metabolite predictors selected on the basis of their capability to discriminate CN versus AD as measured by the Relief algorithm [9] applied on training data in combination with 500 permutations of the outcome variables' values. This method computes the predictors' importance defined as the standardized Relief score, according to Measuring Predictor Importance chapter of [10]. Part of the prediction modeling methodology in this study was adapted after [11], with different algorithms, and followed recommendations from [10,12]. The analysis was carried out using R software [13]. Pathway analysis was performed on the top 20 ranked metabolites using MetaboAnalyst 4.0 [14]. The algorithms were run on four servers with 6-core Xeon CPUs and 336 GB RAM.

3. Results

In this study, we analyzed metabolite data derived from blood samples from 357 participants (CN n = 242, AD n = 115) previously reported in Kim et al. [7]. Demographic and clinical data can be found in [7]; in short, there was no difference in gender while AD participants were older when compared with CN participants.

On the test data, the DL model produced a Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) value of 0.85 with its 95% confidence interval (CI) ranging between 0.8038 and 0.8895. The XGBoost model produced the AUC value of 0.88 (95% CI [0.8619, 0.8903]). When the classifier model RF was employed, the resulting AUC was 0.85 (95%CI [0.8323, 0.8659]). Fig. 1 illustrates ROC curves obtained from the three ML models.

The MC simulation conducted with XGBoost, which was the superior predictive model in our analysis, led to a Gaussian distribution of the AUC values according to [11] and as confirmed by Shapiro-Wilk test (P value = .6819). The 50 AUC values obtained in MC had a minimum of 0.8614, a maximum of 0.8923, a mean of 0.8761, a median of 0.8766 and a standard deviation of 0.0072. The t-test showed that the true mean of AUC for XGBoost applied on plasma metabolites was not lower than 0.87 (P value = 1.265×10^{-07}).

For comparison, we also investigated the levels of amyloid, p-tau and t-tau, to which we added also age and gender, and their prediction for clinical AD versus CN. XGBoost models were built in the same manner as for metabolite predictors. Together with age and gender,
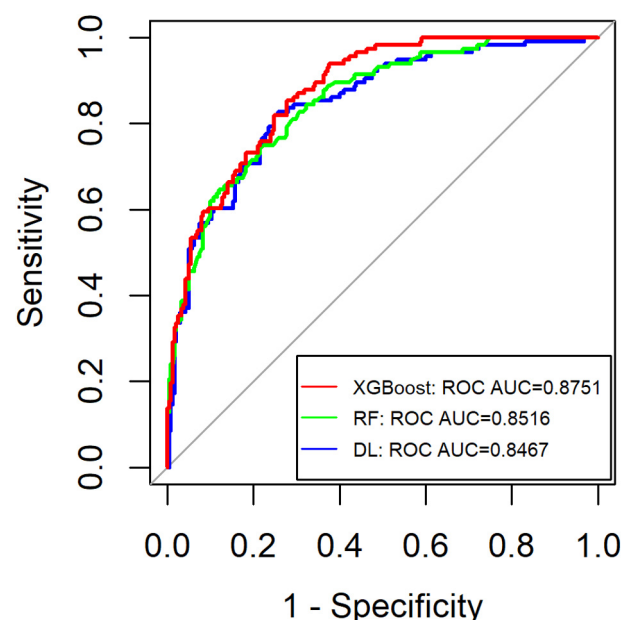

Fig. 1. Shows the AUC values for the XGBoost, RF and DL models. XGBoost performed best with metabolite predictors in the EMIF cohort.

amyloid led to AUC 0.78 (95%CI [0.7626, 0.8013]); p-tau led to AUC 0.83 (95%CI [0.8188, 0.8470]); and t-tau led to AUC 0.87 (95%CI [0.8583, 0.8854]). From the mean AUC for metabolites and for amyloid, p-tau and t-tau calculated individually, the t-tests showed superior values for metabolites (P value$<$2.2 $\times$ 10^{-16}, P value$<$2.2 $\times$ 10^{-16} and P value = .005921, respectively).

The top 20 ranked predictors out of the 347 selected by the method presented in the previous section are shown in Fig. 2.

Pathway analyses revealed that the Nitrogen pathway was overrepresented (q_{FDR} = 0.004) within the panel. Molecules that were captured as the 20 top ranking predictors are discussed in the next section.

4. Discussion

Machine Learning applied to healthcare is increasingly enabled by the advent of high-performance computing and the development of complex algorithms. In this study, we employed two state-of-the-art algorithms, DL and XGBoost, and a more conventional algorithm, RF, to obtain high accuracy models to predict AD versus CN with metabolites as predictors. Our study showed that the best model was based on XGBoost [15], which is an enhanced form of Gradient Boosting Machines methods based on decision trees [12]. In our study RF and DL achieved comparable

AUC. DL algorithms are known to often take advantage of large and/or unstructured data (such as images) to produce more accurate category discrimination/prediction. In a study using the Alzheimer's Disease Neuroimaging Initiative (ADNI) data for AD prediction, XGBoost demonstrated superior results (AUC = 0.97 (0.01) when including imaging parameters (MRI and PET) as predictors and when compared to RF, Support Vector Machines, Gaussian Processes and Stochastic Gradient Boosting [16]. In another study where cognition and MRI were used as predictors, Kernel Ridge Regression was performed to R^2 = 0.87 (0.025) when cognition and MRI predictors were included [17].

Pathway analyses using the top 20 AD predicting metabolites derived from the Relief method showed that the nitrogen pathway was overrepresented. Some of the molecules selected have been reported in metabolomics studies and have been implicated in neurodegeneration: dodecanoate, which is a C12 fatty acid, was found correlated to longitudinal measures of cognition in the ADNI cohort [3] and so was the bile acid glycolithocholate, which was associated to both AD and cognition measures (ADAS-Cog13) in one of the biggest cross-sectional studies on cognition, AD and the microbiome [18]. Plasmalogens were also found in decreased levels in our cohort in agreement with an earlier report [19]. The amide form of vitamin B3, nicotinamide, has been implicated in both neuroprotection and neuronal death [20].
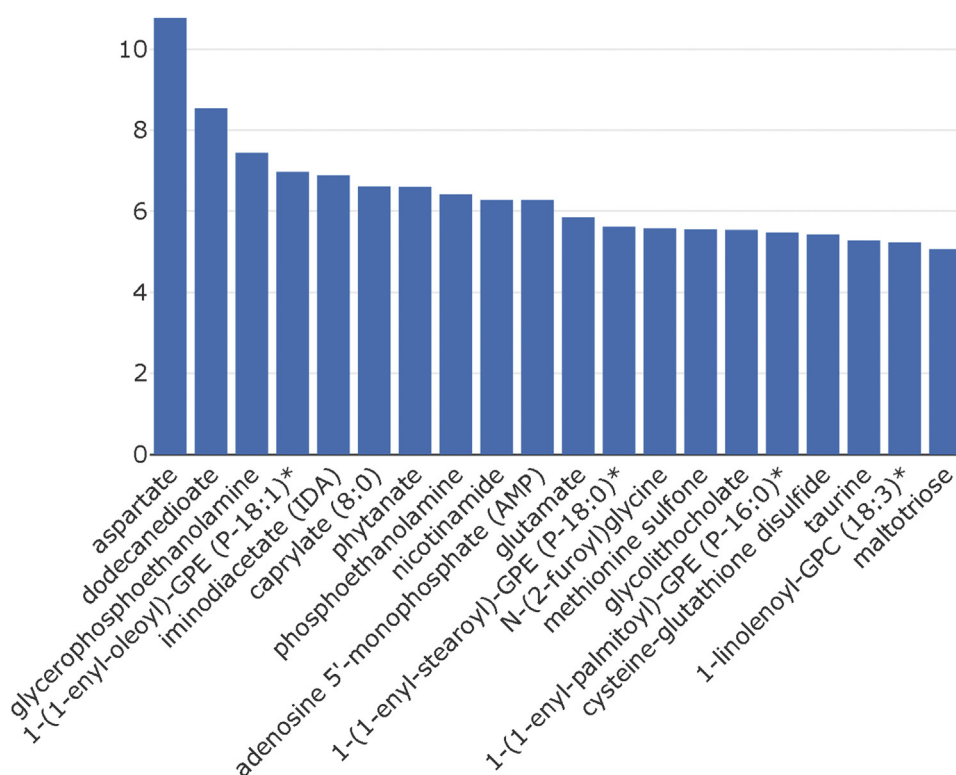

Fig. 2. The *x*-axis shows the top 20 ranked predictors, and the *y*-axis shows the predictors' importance computed as the standardized relief score according to Measuring Predictor Importance chapter of [10].

New metabolites that could be of interest and have not been previously reported as related to AD were phytanate and furoylglycine. The former is a known neurotoxin which impairs mitochondrial function and transcription [21]. Furoylglycine is a metabolite which, as lithocholic acid, is mainly synthesized by the microbiome and has been reported as a biomarker of coffee consumption [22].

A limitation of our study is that it does not include an external validation due to the size of the cohort. However, we implemented a NCV procedure repeated 50 times in a MC simulation that led to an extended internal validation with prediction accuracy of cases. Further studies will assess the performance of ratios/combinations of CSF markers and metabolites, life-style factors and disorders commonly found in the elderly, together with testing the specificity for this specific panel in other neurodegenerative (e.g., PD, FTD), neurological (e.g., stroke) and psychiatric (e.g., depression) disorders associated with aging.

The intent of this paper was to compare the performance of different ML algorithms to identify people with AD from cognitively unimpaired individuals. Here we show first that all three approaches used demonstrate good discriminatory power, second that XGBoost is somewhat more effective in this particular dataset than RF and DL and third, that this accuracy for clinical diagnosis is broadly similar to that achieved by CSF markers of AD pathology. The lack of a replication and validation dataset limits the interpretation of this finding, but nonetheless, the strong prediction of diagnostic category from a blood-based metabolite biomarker set is further evidence of the potential of such approaches to complement other biomarkers in identification of people with likely AD.

Acknowledgments

RESEARCH IN CONTEXT

1 Systematic review: The authors reviewed the literature using *PubMed* and reported key publications. Most AD biomarker studies employing state-of-the-art machine learning (ML) techniques utilized neuroimaging data. Those which looked at blood metabolomics data were small and used clinical diagnosis as an endpoint. Subsequently, we explored the potential of state-of-the art ML algorithms including Deep Learning and Extreme Gradient Boosting to test the performance of blood metabolite levels to clinical diagnosis and compared to CSF biomarkers.

2 Interpretation: The results in here show that with state-of-the-art ML algorithms, blood metabolites have the potential to match the CSF markers of AD pathology on identifying people with AD from cognitively unimpaired individuals. All the ML algorithms employed showed good discriminatory power.

3 Future directions: Results of this study should be replicated and validated using an independent dataset. Further studies will also aim to assess the performance of ratios/combinations of CSF markers and metabolites, life-style factors and disorders commonly found in the elderly, together with testing the specificity for this specific panel in other neurodegenerative (e.g., PD, FTD), neurological (e.g., stroke) and psychiatric (e.g., depression) disorders associated with aging.

References

[1] Proitsi P, Kim M, Whiley L, Pritchard M, Leung R, Soininen H, et al. Plasma lipidomics analysis finds long chain cholesteryl esters to be associated with Alzheimer's disease. Transl Psychiatry 2015;5:e494.

[2] Proitsi P, Kim M, Whiley L, Simmons A, Sattlecker M, Velayudhan L, et al. Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. Alzheimers Dement 2017;13:140–51.

[3] Toledo JB, Arnold M, Kastenmuller G, Chang R, Baillie RA, Han X, et al. Metabolic network failures in Alzheimer's disease: A biochemical road map. Alzheimers Dement 2017;13:965–84.

[4] Varma VR, Oommen AM, Varma S, Casanova R, An Y, Andrews RM, et al. Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. Plos Med 2018;15:e1002482.

[5] Fiehn O. Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol 2002;48:155–71.

[6] Whiley L, Legido-Quigley C. Current strategies in the discovery of small-molecule biomarkers for Alzheimer's disease. Bioanalysis 2011;3:1121–42.

[7] Kim M, Snowden S, Suvitaival T, Ali A, Merkler DJ, Ahmad T, et al. Primary fatty amides in plasma associated with brain amyloid burden, hippocampal volume, and memory in the European Medical Information Framework for Alzheimer's Disease biomarker discovery cohort. Alzheimer's Demen 2019;15:817–27.

[8] Bos I, Vos S, Vandenberghe R, Scheltens P, Engelborghs S, Frisoni G, et al. The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. Alzheimers Res Ther 2018;10:64.

[9] Kononenko I. Estimating attributes: Analysis and extensions of RELIEF. Berlin, Heidelberg: Springer Berlin Heidelberg; 1994. p. 171–82.

[10] Kuhn M, Johnson K. Applied predictive modeling. Springer; 2013.

[11] Stamate D, Katrinecz A, Stahl D, Verhagen SJW, Delespaul P, van Os J, et al. Identifying psychosis spectrum disorder from experience sampling data using machine learning approaches. Schizophr Res 2019;209:156–63.

[12] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. 2 ed. New York: Springer-Verlag; 2009.

[13] Team RC. R: A Language and Environment for Statistical Computing; 2019.

[14] Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic Acids Res 2018;46:W486–94.

[15] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: ACM; 2016. p. 785–94.

[16] Stamate D, Alghamdi W, Ogg J, Hoile R, Murtagh F. A Machine Learning Framework for Predicting Dementia and Mild Cognitive Impairment. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA); 2018. p. 671–8.

[17] Bucholc M, Ding X, Wang H, Glass DH, Wang H, Prasad G, et al. A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. Expert Syst Appl 2019;130:157–71.

[18] MahmoudianDehkordi S, Arnold M, Nho K, Ahmad S, Jia W, Xie G, et al. Altered bile acid profile associates with cognitive impairment in Alzheimer's disease-An emerging role for gut microbiome. Alzheimers Dement 2019;15:76–92.

[19] Han X, Holtzman DM, McKeel DW Jr. Plasmalogen deficiency in early Alzheimer's disease subjects and in animal models: molecular characterization using electrospray ionization mass spectrometry. J Neurochem 2001;77:1168–80.

[20] Fricker RA, Green EL, Jenkins SI, Griffin SM. The Influence of Nicotinamide on Health and Disease in the Central Nervous System. Int J Tryptophan Res 2018;11:1178646918776658.

[21] Schonfeld P, Reiser G. Brain Lipotoxicity of Phytanic Acid and Very Long-chain Fatty Acids. Harmful Cellular/Mitochondrial Activities in Refsum Disease and X-Linked Adrenoleukodystrophy. Aging Dis 2016;7:136–49.

[22] Radjursoga M, Karlsson GB, Lindqvist HM, Pedersen A, Persson C, Pinto RC, et al. Metabolic profiles from two different breakfast meals characterized by (1)H NMR-based metabolomics. Food Chem 2017; 231:267–74.