

---

# Mechanistic links between ApoE4 dose and serum metabolome in Alzheimer's disease

A data science approach

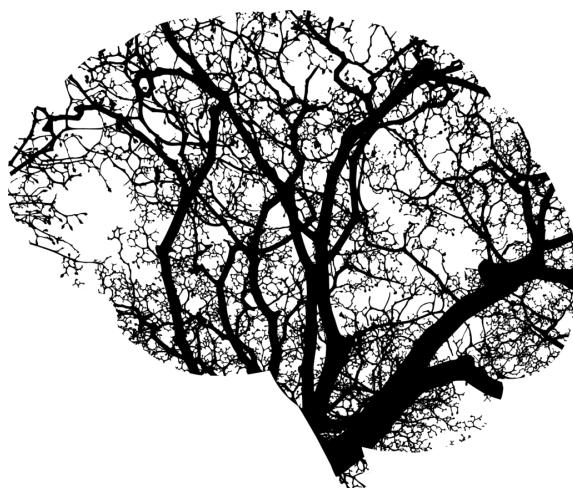
---

George Miliarakis

MSc Thesis  
Wageningen University & Research  
Wageningen, 18 March 2024

*Supervisor*  
**C.F.W. Peeters**  
Mathematical & Statistical Methods (Biometris)  
Wageningen University & Research

*Supervisor*  
**Yannick Vermeiren**  
Nutritional Biology, Human Nutrition & Health  
Wageningen University & Research



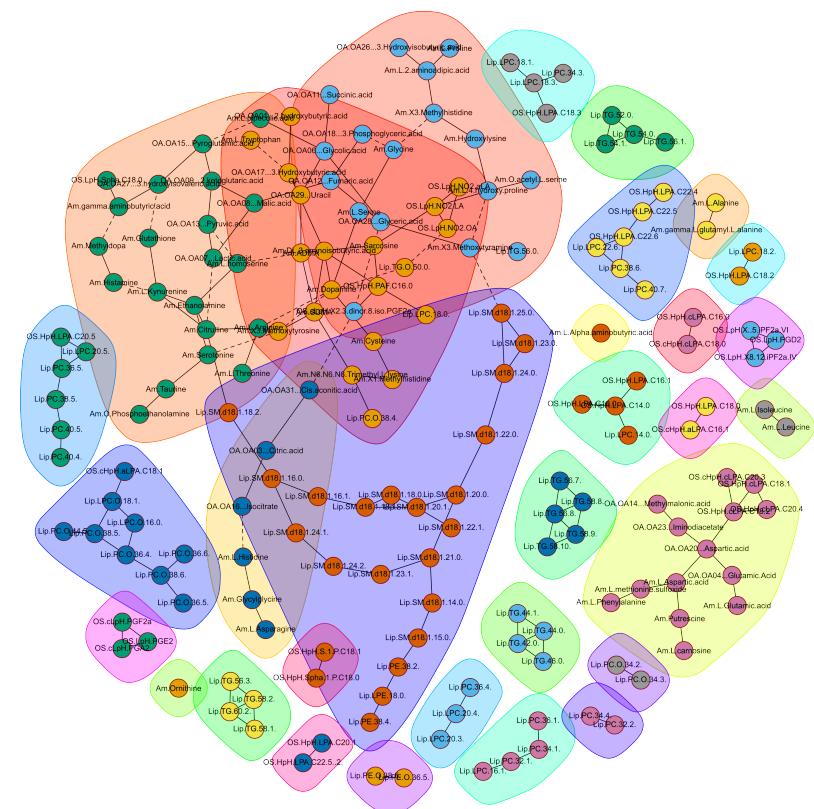
## FOREWORD

This report represents the culmination of my thesis journey, as part of the M.Sc. programme Data Science for Food and Health at Wageningen University & Research. Throughout this journey I acquired a deeper understanding of Alzheimer's disease pathology, as well as several new data science skills (network analysis in R, scientific writing in L<sup>A</sup>T<sub>E</sub>X among others). Given the several scientific disciplines brought together in this project, I regret that this report requires a strong background in human pathophysiology, biochemistry, statistics and machine learning to be comprehended.

I am incredibly grateful to my supervisors, Dr. Carel F.W. Peeters and Dr. Yannick Vermeiren. Their guidance proved instrumental throughout this project. The depth and breadth of their combined academic and scientific expertise fostered an unparalleled learning environment. Dr. Peeters' brilliance was evident in the insightful feedback he always provided and the packages *rags2ridges* and *FMradio* he developed, that streamlined my analyses. My appreciation for Dr. Vermeiren's indispensable contributions is equally profound. His expertise in AD, coupled with his consistently constructive feedback, played a pivotal role in shaping this project. The latter would not have been possible without the invaluable data provided by the Amsterdam Dementia Cohort (confidentiality maintained).



In loving memory of a cherished one who struggled with Alzheimer's disease (AD). Witnessing their struggle (and ultimate passing a month before I embarked on this project) ignited a profound desire in me to contribute to AD research and advocate for improved dementia care.



## ABSTRACT

**Background** Alzheimer's disease (AD), the most common form of dementia, represents a significant global healthcare challenge. Apolipoprotein- $\epsilon$ 4 (ApoE4) is the strongest genetic risk factor for late-onset AD, with nuanced impacts based on sex and ancestry. It is reported that ApoE4 is not metabolized and recycled adequately in the brain, and over time induces the deposition of amyloid- $\beta$  and hyperphosphorylated tau proteins, impairs mitochondrial glucose metabolism and promotes neuroinflammation. There is limited evidence, however, on the effects of ApoE4 on peripheral metabolism. This study explored mechanistic links between the ApoE4 haplotype and serum metabolome in AD and subjective cognitive decline (SCD) patients via three approaches: ApoE4 dose-effects (0, 1, 2 alleles) on individual metabolite levels, multi-class classification of ApoE4 status and AD (4 possible phenotypes: AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4), as well as metabolite precision network analysis among ApoE4 carriers and non-carriers in AD.

**Methods** To screen for ApoE4 dose-effects on serum metabolites, two methods were applied: a global test (correcting only for sex) and nested linear model comparison using ANCOVA F-tests (correcting for several confounding clinical factor), both adjusting for multiple testing via FDR. To test the (added) classification potential of serum metabolites against ApoE4 and AD status, several multi-class classification models (multinomial logistic regression, decision tree and eXtreme gradient boosting) were fitted considering the bias-variance trade-off and model interpretability. Model performance was evaluated via confusion matrices, ROC curves and AUC, as well as accuracy and no-information rate obtained from repeated (100 times) 10-fold CV. To visualize and compare the metabolite (precision) network topologies among ApoE4 carriers and non-carriers, the high-dimensional matrices were first regularized with a Ridge penalty, sparsified and then pruned at a local FDR threshold of 0.001. Network topologies were plotted in Gaussian graphical models (GGMs) and communities of conditionally dependent metabolites were identified using the Girvan-Newman algorithm. The centrality measures were calculated and compared between the ApoE4 carriers and non-carriers using Wilcoxon Signed Rank test.

**Results** In the analysis 120 AD patients and 127 SCD individuals ( $n = 247$ ) were included from the Amsterdam Dementia Cohort. The global test revealed significantly higher tri- and diglyceride levels in AD, while no significant nuances were observed in the SCD group. After adjusting for FDR, no significant ApoE4 dose effects were observed via nested linear model comparison in AD or SCD. However, consistent with the global test, the AD group presented a more pronounced lipid deregulated signature, while the SCD group mainly aminic. Serum metabolites slightly (~ 0.02% AUC) improved model performance and all models had higher accuracy compared to the no-information rate. Putrescine and sphingosine-1-phosphate (oxidative stress compound) emerged as top metabolites in predicting ApoE4 and AD status. The metabolite network of ApoE4 carriers was less random and had higher centrality measures, compared to the non-carriers'. Lastly, the first network could be separated in communities of conditionally dependent metabolites, while the latter was largely more sparse.

**Discussion** These findings suggest beyond perturbed triglyceride metabolism, pronounced aminic and oxidative stress signatures at systematic level associated with ApoE4 presence and (increasing) dose in AD. Further, impaired glucose metabolism (ketoglutarate), increased kynurenine pathway catabolism of tryptophan were observed in AD and SCD, respectively. By incorporating additional "omics" data over time, researchers can potentially develop more sensitive disease-predicting models and stage-dependent biomarkers through large-scale clinical studies. Future investigations could also explore the ApoE4 lipoprotein levels in cerebrospinal fluid (CSF) and serum across the AD continuum, considering as many confounding factors as possible.

**Conclusion** Peripheral metabolism is reported shifted by (the presence or dose) of ApoE4 in AD, considering individual metabolite levels, an ApoE4 and AD status classification signature and network analysis. The proposed methodology paves the way for more elaborate and efficient "omics" data analysis in AD research, potentially leading to the development of improved diagnosis support tools and therapeutic strategies.

**Keywords:** Alzheimer's disease, ApoE4, metabolomics, machine learning, network analysis

## ABBREVIATIONS

- ABCA1:** ATP-binding cassette transporter A1.
- AD:** Alzheimer's disease.
- ApoE:** apolipoprotein E.
- ATP:** adenosine triphosphate.
- AUC:** area under the ROC curve.
- BBB:** blood-brain barrier.
- BMI:** body mass index.
- CSF:** cerebrospinal fluid.
- CV:** cross-validation.
- DT:** decision tree.
- FAD:** familial Alzheimer's disease.
- GC-MS:** gas chromatography tandem mass spectrometry.
- GGM:** Gaussian graphical model.
- HDL:** high density lipoprotein.
- IDL:** intermediate density lipoprotein.
- ISF:** interstitial fluid.
- LOAD:** late-onset Alzheimer's disease.
- LRP1:** low density lipoprotein receptor-related protein 1.
- MAP:** mean arterial pressure.
- MCI:** mild cognitive impairment.
- ML:** maximum likelihood.
- MLR:** multinomial logistic regression.
- NFKB:** nuclear factor kappa B.
- NIR:** no-information rate.
- ROC:** receiver-operating characteristics.
- ROS:** reactive oxygen species.
- RSS:** residual sum of squares.
- SCD:** subjective cognitive decline.
- SMOTE:** synthetic minority over-sampling technique.
- SSR:** sum of squares of the regression.
- TLR4:** toll-like receptor 4.
- UPLC-ESI-Q-TOF-MS:** ultra high-performance liquid chromatography electrospray tandem quadrupole time-of-flight mass spectrometry.
- UPLC-MS/MS:** ultra high-performance liquid chromatography tandem mass spectrometry.
- VLDL:** very low density lipoprotein.
- XGBoost:** eXtreme gradient boosting.

## CONTENTS

Foreword	i
Abstract	ii
Abbreviations	iii
1. Introduction	1
1.1. Alzheimer's disease	1
1.2. Human apolipoprotein-E gene	1
1.3. Human Apolipoprotein-E	3
1.4. ApoE4-mediated metabolic changes in Alzheimer's disease	5
1.5. Research Questions	7
1.6. Approach and Overview	7
2. Methods	8
2.1. Subjects	8
2.2. Data management and software	9
2.3. Outcome measures	9
2.4. Statistical Analysis	10
3. Results	16
3.1. ApoE4 dose effects on serum metabolite levels in AD	17
3.2. Classification of ApoE4 and AD status	20
3.3. Metabolite Network Analysis	23
4. Discussion	27
4.1. Mechanistic links between ApoE4 dose and serum metabolome in AD	27
4.2. Strengths and Limitations	28
4.3. Future directions	29
5. Conclusion	30
References	31
Appendix A. Code and Results	37
Appendix B. Multiclass ROC curves	37
Appendix C. R Session Information	39

## 1. INTRODUCTION

**1.1. Alzheimer's disease.** Alzheimer's disease (AD) is a complex, progressive neurodegenerative disorder and the most common form of dementia [1]. It was considered the 6th leading cause of death in the US in 2019, with an overall increase of 145% in mortality from 2000 to 2019 [2]. The impact AD has on patients, their caregivers, and healthcare systems is detrimental. Hence, a considerable amount of research has been performed in an effort to understand, prevent, impede or cure it. Nonetheless, important aspects of its systematic manifestations remain unknown.

Age is the main risk factor for AD, while several genetic and lifestyle risk factors, as well as biochemical pathways contribute to its development [1]. AD manifests in various histopathological phenotypes and presents a broad spectrum of clinical signs and symptoms [3, 4]. The AD continuum starts with subjective cognitive decline (SCD), followed by mild cognitive impairment (MCI) [5], and continues with progressive loss of global cognition, of which particularly memory, processing speed and executive functioning, spanning a total period of 10-15 years [6].

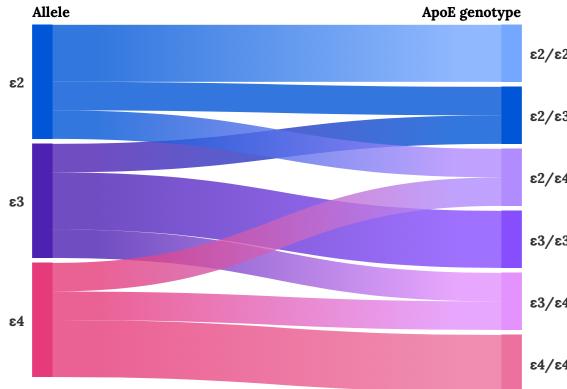
Sporadic or late-onset AD (LOAD; 95% of cases) is the most frequent phenotype, typically appearing after 65 years of age [7]. A rarer phenotype (~ 1%) is early onset familial AD (FAD), usually starting at ages 30–65 and passed in an autosomal dominant fashion [8].

**1.1.1. Amyloid cascade hypothesis.** The amyloid cascade hypothesis has dominated the scientific discussion on the pathogenesis of AD. In historical terms, its impact was profound, as it helped distinguish and identify AD as a singular disease that may be studied for treatment [9]. It suggests that chronic neuroinflammation promotes protein misfolding and accumulation in the brain, forming plaques (consisting of oligomerized amyloid  $A\beta_{42}$ ) and tangles (consisting of hyperphosphorylated tau protein) [4]. Nevertheless, it does not necessarily cover all AD cases. Clinical trials with  $A\beta$  antibodies as a potential disease-modifying treatment have so far failed to demonstrate any clinical improvement in patients, despite successful removal of plaques from AD brains [10, 11]. Evidence suggests that oxidative stress, metabolic abnormalities, atherosclerosis, cardiovascular effects, imbalances of intra-neuronal calcium and metal ion deposition might also contribute to the development of AD [10]. Kepp *et al.* propose a more complex and holistic view of AD pathology, by integrating (epi)genetic, environmental, vascular, neuro-inflammatory and metabolic factors in disease-predicting models [10].

**1.2. Human apolipoprotein-E gene.** LOAD has been associated with various genetic risk factors, but primarily with the gene that encodes for apolipoprotein E (ApoE) [12]. The structure and function of ApoE, as well as how the first defines the latter is described in Section 1.3.

**1.2.1. ApoE4 and Alzheimer's disease.** Humans due to two point mutations in the ApoE gene (rs7412 C/T, rs429358 C/T) present three common variants:  $\varepsilon_2$ ,  $\varepsilon_3$  and  $\varepsilon_4$  [13, 14]. The three ApoE haplotypes are phased in six genotypes [Fig. 1]. In Caucasian populations, the most abundant allele is ApoE3 (rs7412 C, rs429358 T), with a frequency of 78%, and is considered neutral regarding the risk of AD [15]. ApoE4 (rs7412 C, rs429358 C) has a frequency of 14% and represents the strongest genetic risk factor for LOAD, with gene dose effects [16]. Conversely, ApoE2 (rs7412 T, rs429358 T) is found in 8% of Caucasian populations, and is associated with a reduced risk of LOAD [15]. The ApoE haplotypes are strongly linked to the primary pathological features of LOAD, namely  $A\beta$  and phosphorylated tau [17]. While the association between ApoE alleles and LOAD risk or protection is observed across diverse ancestral backgrounds, the strength of this association varies [18, 19], see Section 1.2.3.

Carrying a single  $\varepsilon_4$  allele implies a risk of developing AD of about 20% [20]. A heterozygous ApoE4 allelic composition (e.g.,  $\varepsilon_4/\varepsilon_x$ ) is associated with approximately 3-4 times increased life-time risk of developing AD, while the homozygous composition ( $\varepsilon_4/\varepsilon_4$ ) is associated with 12 times increased risk [21]. Even though ApoE4 is the strongest known genetic risk factor for LOAD, it is neither necessary nor sufficient to cause AD, and it is certainly not the only genetic risk factor [22].



**FIGURE 1.** Sankey chart representing the contribution of the three common ApoE alleles in forming the  $\binom{3}{2} = 6$  ApoE genotypes. Heterozygotes ( $\epsilon_2/\epsilon_3$ ,  $\epsilon_2/\epsilon_4$  and  $\epsilon_3/\epsilon_4$ ) carry two distinct alleles, while homozygotes ( $\epsilon_2/\epsilon_2$ ,  $\epsilon_3/\epsilon_3$  and  $\epsilon_4/\epsilon_4$ ) two copies of the same allele.

**1.2.2. *ApoE and sex synergy in Alzheimer's disease.*** Notably, sex (60% females) and ApoE4 allelic composition (50% has at least one  $\epsilon_4$  allele) are the strongest genetic risk factors for LOAD [23]. In this regard, it is shown that the ApoE4 genotype has a larger impact on females, as they present greater impairment of mitochondrial energy production, compared to males [23, 24].

**1.2.3. *ApoE and ancestry in Alzheimer's disease.*** The majority of studies exploring the relationship between ApoE polymorphism and the genetic aspect of LOAD have primarily focused on Northern European populations[14]. However, smaller studies involving diverse ancestral backgrounds show variations in ApoE4 allele frequencies [14]. While ApoE4 is present in 14% of Caucasian Americans, its prevalence increases to 40% among African Americans, 37% in Oceania, and 26% in Australia. Southern Asia and Europe exhibit ApoE4 allelic frequencies of less than 10%, compared to Northern Europe where it rises to 25% [18, 25–27].

ApoE alleles have different impact on distinct populations. ApoE4 implies a higher risk of LOAD for Korean, Japanese, and Japanese-American, compared to Caucasian populations [19]. Conversely, ApoE4 is associated with a lower risk of LOAD in Native Americans, Hispanic Americans, African Americans, and those of African descent, compared to Caucasian-American populations [19, 28–31]. Notably, ApoE3 is more protective than ApoE2 against AD, as reported by a study in a Chinese population [32]. Some of these population-specific effects are attributed to the ApoE haplotype and ancestral variations in the ApoE gene beyond the  $\epsilon_2/\epsilon_3/\epsilon_4$  haplotypes [28, 30].

**1.2.4. *Evolution of ApoE over time.*** Interestingly, humans are the only species exhibiting polymorphism in the ApoE gene [24]. All other animal species have one ApoE variant, which resembles the human ApoE3 allele [33]. Nonetheless, ApoE4 is the oldest human allele, followed by ApoE3 and ApoE2 in age [24]. ApoE4 may be adaptive by reducing mortality in highly infectious environments, with food scarcity and shorter lifespans [34]. However, as human environments became less septic and provided abundance of food, thus extending life expectancies, ApoE4 started to burden the arteries and brain, increasing the risk of diseases related to ageing [24]. The emergence of ApoE3 from ApoE4 putatively reflects the shift in human diet from a plant-based one to a meat-rich one, where genes adaptive to high meat consumption were, and still are vital to regulate increased cholesterol levels [35].

**1.2.5. *Tissue expression.*** The principal producers of ApoE are hepatocytes in the liver [36]. In the CNS, ApoE is primarily expressed in glia (non-neuronal cells of the brain that support the neurons) and particularly

in astrocytes (metabolic homeostasis and neuronal communication modulators), followed by microglia (the brain immune cells) [37]. Each genotype is linked to different expression levels [13]. ApoE2 carriers seem to have higher cerebrospinal fluid (CSF) levels of ApoE, compared to ApoE4 carriers [38, 39].

### 1.3. Human Apolipoprotein-E.

**1.3.1. Structure and function.** ApoE is a brain-specific lipid-binding glycoprotein of 299 amino-acids (34 kDa, 1 dalton is the mass of 1/12 of C<sub>12</sub>) that comprises several types of lipoproteins, i.e., chilomicra, intermediate density lipoprotein and very low density lipoprotein [13]. Its main function in the brain is the transport of lipids (mainly cholesterol) through membrane receptors [14]. ApoE isoforms have an effect on diverse cellular functions, e.g. synaptic integrity, glucose metabolism, A<sub>β</sub> clearance, blood-brain barrier integrity and mitochondrial regulation [13], see Fig. 3 and 4. How these relate to AD pathology will be elaborated in Section 1.3.4.

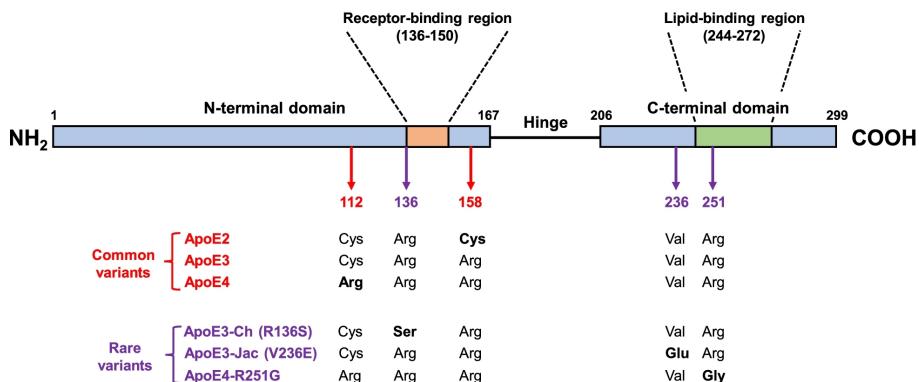
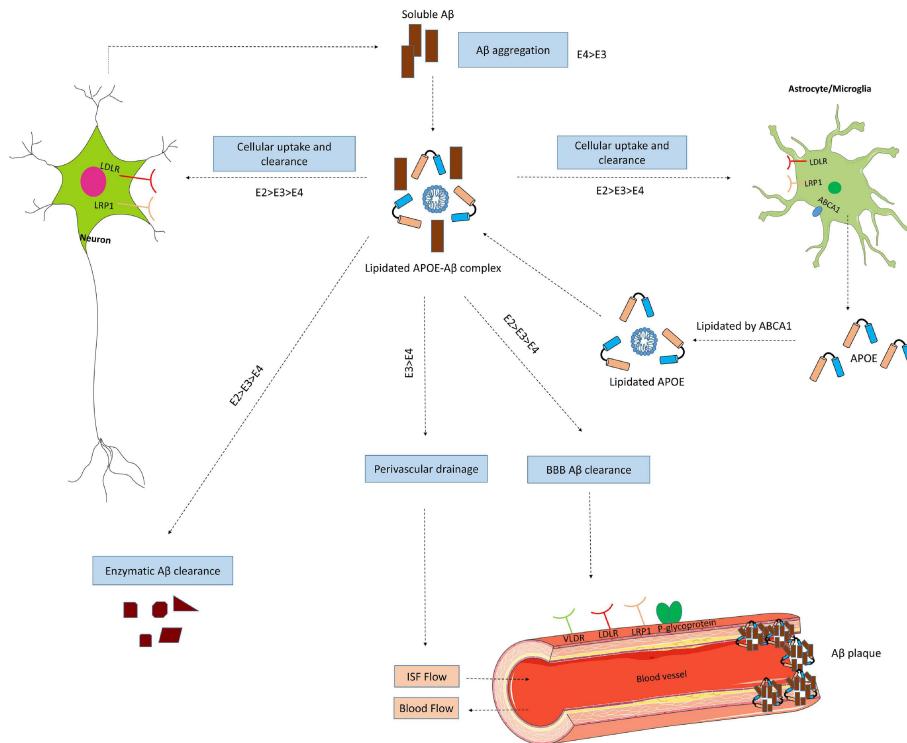


FIGURE 2. Linear representation of the ApoE protein. Three structural domains are highlighted: N-terminal, hinge and C-terminal domains. The different amino acids at positions 112 and 158 are shown for the common alleles and amino acids at positions 136, 236 and 251 encoded by rarer alleles. Source: “APOE targeting strategy in Alzheimer’s disease: lessons learned from protective variants”, Bu (2022) [40]

**1.3.2. ApoE isoforms.** The nuances in the amino acid composition of ApoE, specifically the presence of cysteine or arginine at positions 112 and 158, significantly change its structure and, thus its binding with lipids and receptors [24]. The most prevalent isoform, ApoE3, features cysteine at position 112 and arginine at position 158 [24], as shown in Fig. 2. ApoE2 has two cysteines, while ApoE4 has two arginines at these positions [24]. The C-terminal domain of ApoE (positions 273–299) is crucial for its lipidation specificity and efficiency [41].

**1.3.3. Lipidation nuances of ApoE isoforms.** For ApoE to exert its effects, it needs to be lipidated [13]. ApoE undergoes lipidation via ATP-binding cassette transporter A1 (ABCA1), a lipid efflux protein [42, 43]. The lipidation degree varies among ApoE isoforms, with ApoE4 exhibiting the least efficient lipidation [41, 44]. This discrepancy in lipidation has been linked to alterations in lipoprotein size and type, in that ApoE4 “prefers” large triglyceride-rich VLDL, while ApoE2 and ApoE3 have a higher affinity for phospholipid-rich HDL particles [45]. The lower affinity of ApoE4 for HDL particles leads to increased levels of unlipidated ApoE, resulting in its aggregation [46]. Moreover, ApoE4 fibrils are more neurotoxic than those of ApoE2 and ApoE3 [47].

Poor lipidation leads to poor ApoE recycling [24] (Fig. 3). The latter favours the entrapment of ABCA1 in endosomes, away from the cell membrane, thereby pooling cholesterol in the cell membrane rather



**FIGURE 3.** Effects of ApoE isoforms on the metabolism and removal of A $\beta$ . In the brain, ApoE is mainly expressed in astrocytes and microglia, and undergoes lipidation by ATP-binding cassette transporter A1 (ABCA1) creating lipoprotein particles. ApoE increases the accumulation and build up of A $\beta$ , or promotes cellular uptake of A $\beta$  by astrocytes or microglia via endocytosis of the lipidated ApoE-A $\beta$  in an isoform-specific manner. This process involves several receptors, such as the low-density lipoprotein receptor and low density lipoprotein receptor-related protein 1 (LRP1). ApoE also facilitates the breakdown of A $\beta$  outside the cells in an isoform-specific manner. At the blood-brain barrier, soluble A $\beta$  is predominantly transported from the interstitial fluid (ISF) into the bloodstream via LRP1 and P-glycoproteins. Additionally, ApoE plays a role in the perivascular drainage of A $\beta$ . Insufficient clearance of A $\beta$  can lead to its accumulation in the brain tissue, contributing to the formation of amyloid plaques. Source: “APOE and Alzheimer’s Disease: From Lipid Transport to Physiopathology and Therapeutics”, Husain *et al.* (2021) [13]

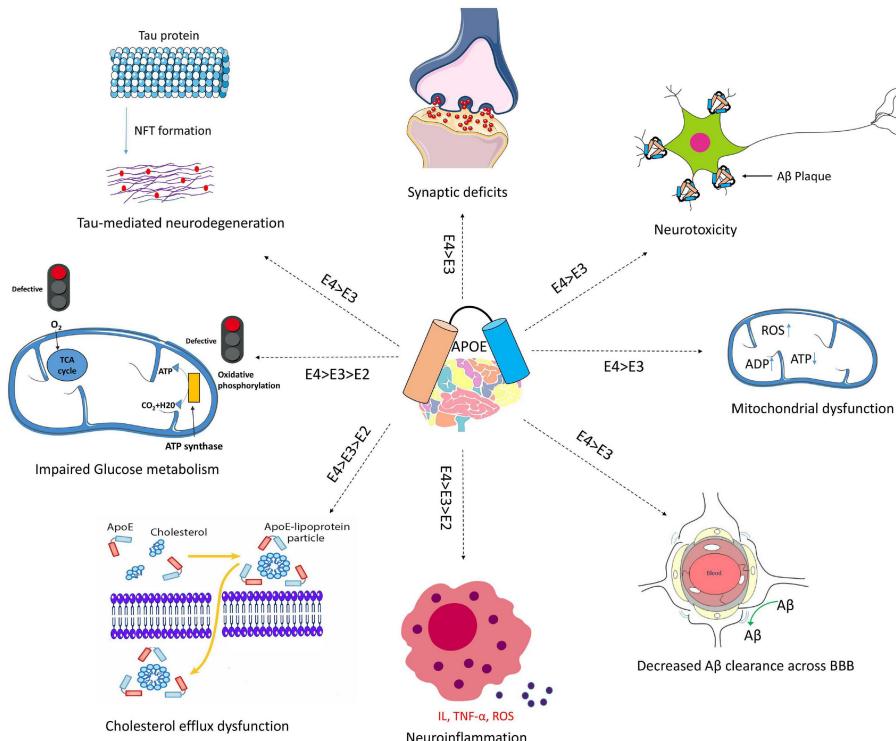
than attaching it to HDL particles [48]. This increased cholesterol content in the cell membrane amplifies toll-like receptor 4 (TLR4) signalling in macrophages, activating NF $\kappa$ B and inducing an inflammatory gene response [24]. ApoE4 accumulation also sequesters the insulin receptor (IR) in endosomes, impacting cellular energy preferences [49]. This leads to a decrease in glucose utilization for ATP production (Fig. 4) and an increase in fatty acid oxidation [50].

**1.3.4. Interplay between ApoE lipidation and Alzheimer’s disease pathophysiology.** As mentioned earlier, ApoE isoforms have differential pleiotropic effects on various cellular functions. ApoE4 induces a pro-inflammatory response, leading to the dysfunction of the blood-brain barrier, which in turn impairs cognitive

functions [51–53]. Moreover, ApoE modulates the primary neuropathological hallmarks of AD: neuroinflammation, A $\beta$  plaques and tau tangles [13]. Evidence from human and transgenic mice studies reveals increased brain A $\beta$  and amyloid plaque loads in ApoE4 carriers, compared to ApoE3; with the lowest levels in ApoE2 carriers [54–56]. ApoE4 has a higher binding affinity for A $\beta$  which leads to impaired clearance, its intracellular aggregation and higher plaque loads [53] (Fig. 3). Additionally, high levels of ApoE4 in neurons remarkably increase tau protein phosphorylation, while high concentrations of ApoE3 do not seem to have an effect [57–60]. Notably, ApoE directly inhibits phosphorylation of tau by glycogen synthase kinase-3 $\beta$  [61]. An overview of the A $\beta$ -(in)dependent effects of ApoE is shown in Fig. 4.

**1.4. ApoE4-mediated metabolic changes in Alzheimer's disease.** Metabolism entails the repertoire of chemical reactions that keep living organisms alive. Metabolites -the molecules involved-, and especially lipid [62–64], perceived as functional intermediates, are rigorously studied as biomarkers or therapeutic targets in AD [65].

**1.4.1. Measured in post-mortem brain tissue.** A metabolomic profiling of brain tissue, obtained *post-mortem* from AD patients and healthy controls revealed pronounced impairments in sterol and sphingolipid levels in ApoE4 carriers with AD [66]. However, another *post-mortem* metabolomic analysis did not reveal

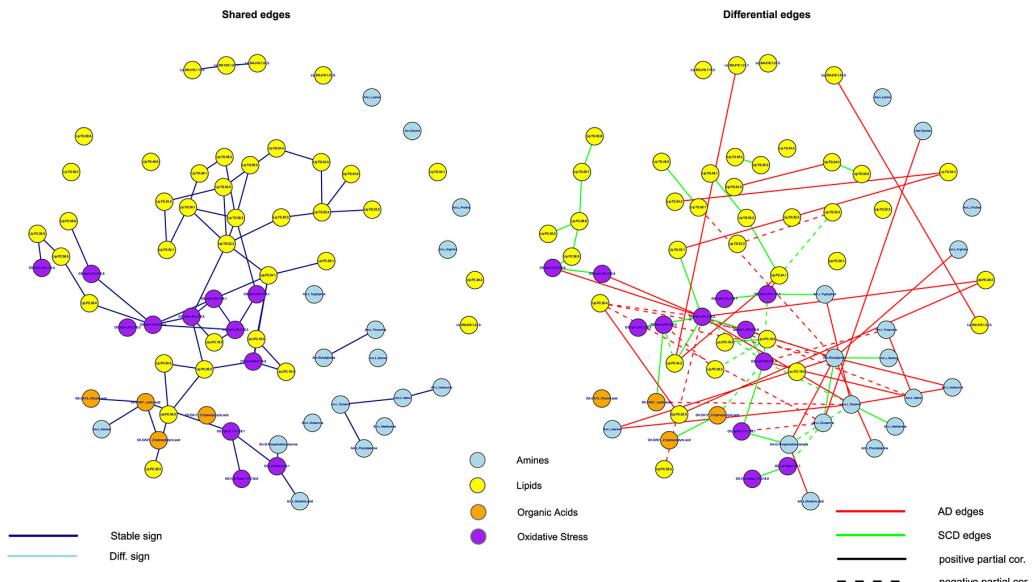


**FIGURE 4.** Schematic overview of A $\beta$ -independent effects of ApoE in AD pathology. ApoE4 increases the phosphorylation of tau proteins, leading to the creation of tangles, inducing neurodegeneration, synaptic deficits and neurotoxicity. Moreover, ApoE4 is associated with decreased cholesterol efflux and neuroinflammation mediated by interleukin, tumour necrosis factor  $\alpha$  and reactive oxygen species (ROS). In the mitochondria, ApoE4 impairs glucose metabolism and ATP production. Source: “APOE and Alzheimer’s Disease: From Lipid Transport to Physiopathology and Therapeutics”, Husain *et al.* (2021) [13].

nuances significantly correlated with ApoE4 [67], although they showed trends of increased cholesterol esters, unsaturated lipids, and sphingomyelin species.

**1.4.2. Measured in blood.** Transcriptomics and lipidomics analyses in humanized ApoE mice associated ApoE4 with decreased free fatty acid levels, many increased tricarboxylic acid (TCA) cycle metabolites, as well as changes in plasma levels of phosphatidylcholines and unsaturated fatty acids [68, 69]. A recent study on 58 individuals found six downregulated plasma metabolites (including lysophospholipids and cardiolipin) in ApoE4 carriers [70]. Further, the plasma metabolome of the latter reveals a preference for aerobic glycolysis [71]. Significant correlations of ApoE genotype and sex with metabolites were observed, i.e. several phosphatidylcholines were found in a large study of more than 1500 individuals [23].

Perturbed serum metabolites associated with AD are aminoacids, amines [72, 73], cholesteryl esters [64], sphingolipids [62, 65, 73–75], fatty acids [63, 72], glycerophospholipids [74, 76–78], phosphatidylcholines [79] and lipid peroxidation compounds [63]. These molecules are usually identified via high-throughput metabolomic pipelines (coupled with Mass Spectrometry (MS) detectors) that trace all compounds in a sample and result in high-dimensional data [80]. The latter often require advanced statistical methods e.g. projection to latent structures [78, 81] or graphical models [82] in order to extract putatively meaningful information. With such techniques, de Leeuw *et al.* discovered distinct serum metabolic signatures among AD patients-controls and those carrying at least one ApoE4 allele [72], as they are shown in Fig. 5. The different intra-group metabolic profiles, however, among ApoE4 carriers and non-carriers remain obscure.



**FIGURE 5.** Mutual (left-hand panel) and distinct (right-hand panel) metabolic network topologies between ApoE4 carriers with AD and non-carriers with SCD, as published by de Leeuw *et al.*. Red edges represent links that are present exclusively in ApoE4 carriers. Green edges represent connections found in the SCD group without ApoE4. Solid edges represent positive partial correlations, while dashed edges represent negative partial correlations. Abbreviations: SCD, subjective cognitive decline. Source: “Blood-based metabolic signatures in Alzheimer’s disease”, de Leeuw *et al.* (2017) [72].

**1.5. Research Questions.** ApoE4 carriers –particularly females– experience metabolic disturbances and are at increased risk of LOAD. The mechanistic links, however, between ApoE4 dose, metabolism and AD development are not entirely known [63]. Tracking ApoE4 dose effects on serum metabolites might reveal metabolic perturbations at systematic level, preceding or concurring with AD. To this end, one may focus on shifted individual metabolite levels, a classification signature, a network signature or all of them. Considering all three approaches, one could state the following overarching research question and subquestions:

- Are there mechanistic links between ApoE4 dose and serum metabolome in AD?
  - i Are there ApoE4 dose effects on serum metabolite levels in AD?
  - ii What is the (added) classification potential of serum metabolites in predicting ApoE4 and AD status?
  - iii How do the metabolite network topologies differ between ApoE4 carriers and non-carriers?

**1.6. Approach and Overview.** To facilitate statistical analysis, two new features were created as outcome measures for the first two research questions: "ApoE4dose" (0, 1, 2) and "target" (4 possible phenotypes: AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4), respectively, as described in Section 2.3. Next, an overview of the approaches taken to address the research subquestions is provided.

*Are there ApoE4 dose effects on serum metabolite levels in AD?* To investigate ApoE4 dose effects on individual metabolites two methods were applied: a global test (correcting only for sex) and nested linear model comparison using ANCOVA F-tests (correcting for several factors: Table 1, except CSF markers).

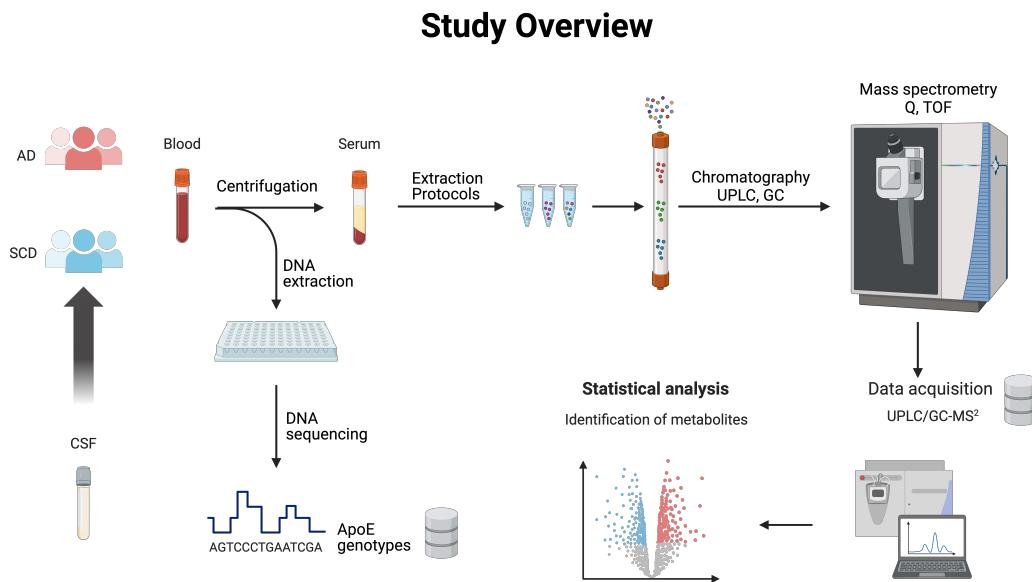
*What is the (added) classification potential of serum metabolites in predicting ApoE4 and AD status?* To test the (added) classification potential of serum metabolites against the target, several multi-class classification models were fitted considering the bias-variance trade-off. First, a benchmark multinomial logistic regression model was fitted using only clinical background data as predictors. Second, the full metabolomic panel was added on top of the clinical data in the same model and feature importance was calculated. Third, the metabolites were projected to a latent orthogonal space, where 6 meta-metabolites (accounting for around 30% of the variance) replaced the original metabolites. Finally, the meta-metabolites were fitted on top of the clinical data in a multinomial logistic regression model, a decision tree and a eXtreme gradient boosting (XGBoost) model. The discriminatory performance of the aforementioned models was comprehensively evaluated and compared.

*How do the metabolite network topologies differ between ApoE4 carriers and non-carriers?* To visualize and compare the (precision) metabolite network topologies among ApoE4 carriers and non-carriers, the high-dimensional matrices were first regularized with a Ridge penalty, sparsified and pruned controlling the False Discovery Rate. The metabolites were plotted in GGMs, their centrality measures were calculated and compared between the two groups using Wilcoxon Signed Rank test.

An introduction to the data used in this study is found in Section 2.1. A general overview of the software is found in Section 2.2, while the R session information is found in Appendix C. The results of the analysis are presented in Section 3 and discussed in Section 4. The study is concluded in Section 5.

## 2. METHODS

**2.1. Subjects.** The data were collected from 120 AD patients and 127 SCD ( $n = 247$  in total) individuals recruited within the Amsterdam Dementia Cohort [72, 83]. The AD diagnosis was confirmed with a ratio of CSF markers  $t\text{-tau}/A\beta_{42} > 0.52$ . In this study two data sets were analysed: a semi-targeted metabolomics panel and clinical background data, i.e. age at diagnosis, sex, smoking status, alcohol consumption, hypertension (and medication), hyperlipidemia (and medication), anticoagulant medication, antidepressants, mean arterial pressure (MAP) and body mass index (BMI) (see Table 1). The metabolomics set contains  $p = 230$  metabolites (amines, organic acids, lipids and oxidative stress compounds) measured in serum obtained from peripheral blood (see Fig. 6). Amines and oxidative stress compounds were measured via ultra high-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS), organic acids with gas chromatography tandem mass spectrometry (GC-MS) and lipids with ultra high-performance liquid chromatography electrospray tandem quadrupole time-of-flight mass spectrometry (UPLC-ESI-Q-TOF-MS). The detailed methodology for the metabolomic analysis and ApoE genotyping can be found at de Leeuw et al.'s Blood-based metabolic signatures in Alzheimer's disease: SMT1 [72]. The data were cleaned as described in the same article. The metabolomics datasets for AD and SCD are high-dimensional, in the sense that they contain more variables than observations ( $p > n$ ). Another particularity of the data is the variable collinearity. Therefore, appropriate measures need to be taken to prevent model over-fitting –the algorithm being unable to distinguish signal from noise and fitting the latter– and to correct for spurious correlations.



**FIGURE 6.** Schematic overview of the study methodology. CSF markers supported the diagnosis of AD or SCD. Peripheral blood was obtained from both groups for a semi-target metabolomics analysis and ApoE genotyping. For the metabolomics, the blood samples were centrifuged, the serum supernatant was obtained and processed distinctly for each metabolite class. Metabolite levels were measured using UPLC/GC-MS<sup>2</sup> and recorded in the data used in this study. The latter pertains to the Statistical Analysis part of the process.

TABLE 1. Clinical background characteristics used as control variables of ApoE4 (dose) effects.

	Clinical feature	Values	Data type
Anthropometric	Age	years	Discrete
	Sex	m, f	Binary
Intoxications	Smoking	past, current, no	Nominal
	Alcohol	yes, no	Binary
Comorbidities	Hypertension	yes, no	Binary
	Diabetes mellitus	yes, no	Binary
	Hypercholesterolemia	yes, no	Binary
Medication	Cholesterol lowering	yes, no	Binary
	Antidepressants	yes, no	Binary
	Antiplatelet	yes, no	Binary
CSF markers*	A $\beta$ <sub>42</sub>	pg/mL	Continuous
	tau	pg/mL	Continuous
	p-tau	pg/mL	Continuous

\* Were only used in the ApoE4 and AD status multi-class classification.

**2.2. Data management and software.** The FAIR principles for data management and stewardship in science were published by Wilkinson *et al.* in 2016 [84]. FAIR stands for Findable, Accessible, Interactive, and Reusable data; the intention is to create and use data that are well-documented and reproducible. These principles were considered at every aspect of the study and implemented when applicable.

The statistical analyses were performed in R (version 4.3.2), and the current report was redacted in L<sup>A</sup>T<sub>E</sub>X (Tex Live version 2023), both in VSCode (the open-source version of VSCode) as integrated development environment (IDE). Vector graphics (e.g. metabolite network analysis plots) were edited using Inkscape. The schematic overview of the study was created in Biorender. All files (except for the data) are stored in a private github repository.

**2.3. Outcome measures.** The information of the ApoE genotypes is valuable, and it might be interesting to screen for metabolic nuances between them. However, the genotypes were not equally represented in the data and hence, testing for differences among them would not be possible. The following two subsections describe the features ApoE4dose and target that were created as outcome measures to balance the genotype counts.

**2.3.1. ApoE4 dose effects.** De Leeuw *et al.* divided the subjects into two classes: those carrying at least one  $\epsilon_4$  allele, and  $\epsilon_4$  non-carriers. In order to study the  $\epsilon_4$  dose effects, the genotypes can be categorized into groups, based on the number of  $\epsilon_4$  alleles they carry: 0, 1 or 2. The number of ApoE4 allele doses is shown in the bottom part of Table 2.

**2.3.2. Classification of ApoE4 and AD status.** In order to incorporate ApoE4 status, as well as the diagnosis of AD, a four-level feature (AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4, "target") was created, as shown in Table 2.

TABLE 2. Observed ApoE genotype (top part), allele (middle part) and ApoE4 allelic (bottom part) counts in AD and SCD

		AD (%)	SCD (%)	Total (%)
<i>genotypes</i>	$\varepsilon_2/\varepsilon_2$	2 (1.7)	0 (0.0)	2 (0.8)
	$\varepsilon_2/\varepsilon_3$	15 (12.5)	3 (2.4)	18 (7.3)
	$\varepsilon_2/\varepsilon_4$	5 (4.2)	2 (1.6)	7 (2.8)
	$\varepsilon_3/\varepsilon_3$	69 (57.5)	37 (29.1)	106 (42.9)
	$\varepsilon_3/\varepsilon_4$	26 (21.7)	59 (46.5)	85 (34.4)
	$\varepsilon_4/\varepsilon_4$	3 (2.5)	26 (20.5)	29 (11.7)
<i>alleles</i>	$\varepsilon_2$	24 (10.0)	5 (2.0)	29 (6.0)
	$\varepsilon_3$	179 (74.6)	136 (53.5)	315 (64.0)
	$\varepsilon_4$	37 (15.4)	113 (44.5)	150 (30.0)
$\varepsilon_4$	1x*	31 (25.8)	61 (48.0)	92 (37.2)
	2x*	3 (2.5)	26 (20.5)	29 (11.7)
	$\geq 1x^\dagger$	34 (28.3)	87 (68.5)	121 (49.0)
	No* $^\dagger$	86 (71.7)	40 (31.5)	126 (51.0)
	Total	120 (49.0)	127 (51.0)	247

\* rows used for ApoE4dose

$^\dagger$  rows used for target

## 2.4. Statistical Analysis.

2.4.1. *ApoE4 dose effects on serum metabolite levels in AD.* To test if the number of ApoE4 alleles have an effect on mean metabolite levels in AD, two methods were applied: a global test and nested linear model comparison with ANCOVA.

**Global Test** The concept of a global test was first introduced by Simon *et al.* to cater to the high dimensionality of gene expression data [85]. The R package `globaltest` features a multinomial logistic regression model, fitting genes to predict clinical or biological group membership [86–88], number of ApoE4 alleles in this case. This method is also appropriate for other types of -omics data, such as metabolomics in this study [88]. The null hypothesis is that metabolite levels are independent of the ApoE4 dose X, i.e.  $H_0 : P(Y|X) = P(Y)$ , where  $X \in \mathbb{R}^{n \times p}$ . The test statistic under  $H_0$  follows, asymptotically, a normal distribution. The `gt` function of the package was used to screen for nuances in metabolite levels on the number of ApoE4 alleles, correcting for sex. To assess ApoE4 dose effects correcting for clinical data nested linear model comparison was performed, as described in the next section.

**Nested linear models** A linear regression model using the least squares principle estimates the sum of squares of the regression (SSR)

$$\text{SSR} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

where  $\bar{y}$  is the sample mean and  $\hat{y}_i$  is the estimate for the  $i$ -th observation, using categorical and numerical variables. It does so while minimizing the residual sum of squares (RSS) [89],

$$\text{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where  $y_i$  is the  $i$ -th observation [89]. Two models were fitted for each metabolite, a full and a nested model. The dependent variable in each model was a metabolite; the nested model (1), has only clinical variables (Table 1 except CSF markers) while the full model (2), features the clinical variables, plus the number of ApoE4 alleles (0, 1 or 2 -nominal) as explanatory variables. Let  $y_j$  represent the  $j$ -th metabolite,  $x_k$  the  $k$ -th clinical variable, and  $D_{e4}$  the ApoE4 dose; the nested linear regression models are:

$$(1) \quad y_j = \beta_0 + \sum_{k=1}^m \beta_k x_k + \epsilon$$

$$(2) \quad y_j = \beta_0 + \sum_{k=1}^m \beta_k x_k + \beta_{m+1} D_{e4} + \epsilon$$

For every metabolite  $j = 1, \dots, p$  the hypothesis test is

$$H_0 : \beta_{m+1} = 0 \text{ vs } H_\alpha : \beta_{m+1} \neq 0$$

The test statistic is ANCOVA F-test:

$$F = \frac{\text{SSR}_{full} - \text{SSR}_{nested}/df_{full} - df_{nested}}{\text{RSS}_{nested}/n - (m + 2)}$$

Under  $H_0$   $F \sim \mathcal{F}_{1,n-(m+2)}$  [89]. The null hypothesis is rejected for large values of  $F$ .

This implies  $p = 230$  hypothesis tests. The probability of incorrectly rejecting  $H_0$  (type I error) increases monotonically with every additional test. Multiple testing provides tools to correct the increase of type I errors, i.e. controlling False Discovery Rate [90]. That is controlling the expected ratio of incorrectly rejected  $H_0$  hypotheses, globally e.g. at a  $q$  of 0.05. With the Benjamini-Hochberg's approach (see Algorithm 1), first, the p-values are sorted in ascending order. Then, for every  $j$  p-value,  $q$  is multiplied by  $j$  over the total number of tests [90]  $m = 230$  in the AD group in this study. In other words, after adjustment, a null hypothesis  $j$  may be rejected only if its associated F-test's p-value is less or equal to a fraction  $(j/m)$  of  $q$ .

---

**Algorithm 1:** Benjamini–Hochberg's procedure to control FDR. Source: [91]

---

- 1 Specify  $q$ , the level at which to control the FDR.
- 2 Compute p-values,  $p_1, \dots, p_m$ , for the  $m$  null hypotheses  $H_{01}, \dots, H_{0m}$ .
- 3 Order the  $m$  p-values so that  $p(1) \leq p(2) \leq \dots \leq p(m)$ .
- 4 Define

$$L = \max \left\{ j : p(j) \leq \frac{j}{m} q \right\}$$

- 5 Reject all null hypotheses  $H_{0j}$  for which  $p_j \leq p(L)$ .
- 

To test for ApoE4 dose effects on each metabolite in AD and SCD,  $2m = 460$  model comparisons were needed in total, hence a function was created to iterate over the metabolites, fit the nested models, compare them using ANCOVA F-tests, store and adjust the p-values, filter those below .05, consolidate the coefficients of the meaningful full models and the p-values of their t-tests in a table and display it. To decrease run time, as well as harness the power of multiple CPU cores, furrr's `future_map` function was used for parallel iterations.

**2.4.2. Classification of ApoE4 and AD status.** In statistical learning, a class denotes a group of objects that share common characteristics, such as having AD or ApoE4 [92]. Classification, in this context, denotes training a (supervised) learning algorithm on labelled data (containing their class) [92]. The classifier learns patterns in the latter and may be able to predict class membership for unknown data [92].

**Bias-Variance trade-off** The degree to which a user can understand and interpret the prediction or decisions made by a statistical model is defined as *interpretability* [93]. It was of interest in this study to find the optimal balance between the performance of a model and its interpretability. The *bias-variance trade-off* was formally introduced by Geman *et al.* and refers to the trade-off between the accuracy (opposite of bias) and precision (opposite of variance) of a prediction. It also refers to the trade-off between model flexibility (or complexity) and interpretability [94]. One may consider this trade-off during model and evaluation method selection, as some impose more bias or variance than others.

**Multi-class classification models** Considering interpretability, Multinomial Logistic Regression (MLR) is inherently interpretable. Let  $y$  a response with  $K$  classes,  $k \in N, [1, 4]$  representing the  $k$ -th class of the target and  $\beta_{kj}$  its set of coefficients,  $\beta_{lj}$  the coefficients of the rest of classes for  $j$ -th metabolite, a multinomial logistic regression model calculated the probability

$$P(y = k|X = x) = \frac{e^{\sum_{j=1}^p \beta_{kj} x_j}}{\sum_{l=1}^K \sum_{j=1}^p e^{\beta_{lj} x_j}}$$

In the  $K$ -way classification, the problem is reduced to  $\binom{K}{2}$  binary classifications with the one-vs-one method. ApoE4 and AD status contains  $K = 4$  classes, so 6 binary classifications were performed [91]. The function `multinom` of the package `nnet` was used to fit a shallow neural network (with a single hidden layer, but allowing skip-layer connections) [95]. For the  $i$ -th observation, it calculates the weights via a Least Squares estimation of the negative conditional log-likelihood that it belongs to the  $k$ -th class

$$E = \sum_i \sum_k -y_{ik} \log \hat{y}_{ik}, \quad \hat{y}_{ik} = \frac{e^{\sum_{j=1}^p \hat{\beta}_{kj} x_{ij}}}{\sum_{l=1}^K \sum_{j=1}^p e^{\hat{\beta}_{lj} x_{ij}}}$$

where  $y_{ik}$ , the true class will be exactly one and the others all zero [95].

When  $p > n$ , it is not possible to perform classification fitting all the predictors [91]. Regularization trades off a small increase in bias for a great decrease in variance, by shrinking the low coefficients towards zero [91]. Ridge or L2 regularization [96] shrinks the coefficients without setting them to exactly zero [96]. The predictors were regularized with weight decay, a Ridge-like penalty based on the sum of squares of the weights [95]:

$$\lambda \sum_{j=1}^p \beta_j^2$$

where  $\lambda \geq 0$  is a tuning parameter that balances the coefficient shrinking effect.

Another method to treat multi-collinearity and high dimensionality is a 2-stage maximum likelihood (ML) Factor Projection to a Latent orthogonal space, such as the one the package `FMradio` [81] performs. In the 1st stage, a ML estimation is used to filter out redundant features from the data matrix. In the second stage, the latter is projected to an orthogonal space where the features are replaced by -fewer- L2-regularised factors explaining their covariance. One can then use the produced factor scores as predictors in any model.

Decision Trees (DT) are inherently interpretable, non-parametric models, that fit well large and complicated data sets. They have a tree-like structure that splits the data based on a threshold into branches and leaves (nodes) [97]. Several studies used decision trees to support AD diagnosis [98–101]. The `rpart2` [102] function of `rpart` was used.

Boosting models, are ensemble models that fit several weak learners (such as linear/logistic regression or DTs) sequentially, reweighing the data, and take their weighted majority vote [103, 104]. Even though Boosting tends to outperform DTs, it often operates as a *black box* and is poorly interpretable. The state-of-the-art eXtreme gradient boosting (`xgbTree`) of the package `xgboost` [105] was used. Ensemble learners, and XGBoost in particular, is also being used to predict multiclass AD status [106, 107].

**Implementation** The R package `caret` streamlined the training and comparison of classification and regression models, offering broad model training and evaluation options, as well as feature importance estimation [108]. The functions `trainControl` and `train` were invoked to fit the models. The sampling method used to handle the unbalanced classes was SMOTE (Synthetic Minority Over-Sampling Technique), as implemented by the package `DMwR2` [109]. Considering the bias-variance trade-off, three models were implemented: multinomial logistic regression (parametric, interpretable), decision tree (non-parametric, sufficiently interpretable), XGBoost (non-parametric, not interpretable).

First, a benchmark model was created fitting the clinical background data to predict the target in an unpenalized MLR model ( $\lambda = 0$ ) (Fig. 7). Second, the 230-metabolite panel was fitted on top of the clinical background data in a penalized MLR model ( $\lambda = 9.187724$ , obtained from repeated 10-fold Cross-Validation (CV)) and feature importance was calculated invoking `caret`'s `varImp`. Next, the full metabolite panel was projected into 6 latent factors (meta-metabolites), explaining around 30% of their variance. The number of latent factors was selected assuming four meta-metabolites account for the variance of the four metabolite classes (amines, lipids, organic acids and oxidative stress compounds) while balancing model complexity and cumulative explained variance. The 6-factor metabolite projection was then fitted on top of the clinical data in a unpenalized MLR ( $\lambda = 0$ ), a decision tree and an XGBoost model. The aforementioned models were hyperparameter-tuned over a grid of values and the best "tune" was selected via repeated (100 times) 10-fold CV. The full metabolite panel was initially fitted in a DT and a XGBoost as well, but was omitted from further runs, as their added value was deemed small considering that more parsimonious models with only 19 predictors (clinical data + 6 meta-metabolites) would be fitted next.

**Evaluation** The models were fitted with the optimal hyperparameters. Model performance was assessed and compared with repeated (100 times) 10-fold CV-obtained receiver-operating characteristics (ROC) curves and their respective area under the ROC curve (AUC) using the `pROC` package [110] and other metrics such as accuracy and no-information rate (NIR) from `caret`'s `confusionMatrix` output. The confusion matrices were also plotted using the R package `cvms`.

ROC plots feature sensitivity (recall, true positive rate)

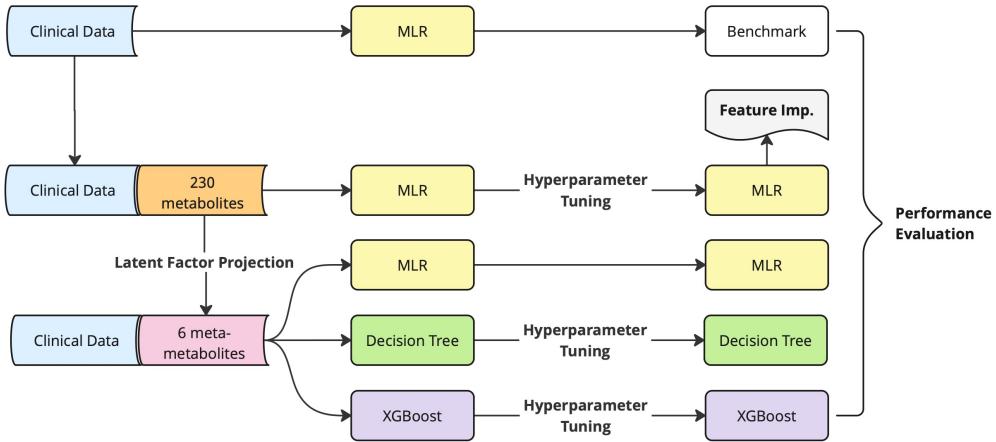
$$sensitivity = \frac{TP}{TP + FN}$$

on the y-axis against 1-specificity (false positive rate)

$$specificity = \frac{TN}{FP + TN}$$

$$1 - specificity = \frac{FP}{FP + TN}$$

on the x-axis [91]. A well-performing classifier has high sensitivity for low false positive rates, resulting in a curve that tends to stay close to the top-left corner of the graph and encloses a large AUC [91]. The latter represents the probability that the classifier will correctly distinguish the classes. An AUC of 1 indicates perfect performance, while an AUC of 0.5 is equivalent to random chance [91].



**FIGURE 7.** Multi-class classification of ApoE4 and AD status pipeline. First, fit a benchmark model: clinical data only in MLR with  $\lambda = 0$ . Second, fit clinical data + 230 metabolites in MLR, hyperparameter tune, use the best  $\lambda$ , evaluate performance and get feature importance scores. Third, project the 230 metabolites to 6 latent factors (metab-metabolites), fit these on top of the clinical data in MLR and evaluate performance. Fourth, fit the same data in a decision tree, hyperparameter tune, use the best hyperparameters and evaluate performance. Fifth, fit the same data in an eXtreme gradient boosting (XGBoost), hyperparameter tune, use the best hyperparameters and evaluate performance. Performance was evaluated for all models using P(Accuracy  $\geq$  NIR), ROC, AUC, confusion matrices obtained from repeated (100 times) 10-fold CV.

MLR: multinomial logistic regression

Accuracy represents the percentage of correct predictions over all predictions:

$$\text{accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

NIR is the expected accuracy of a classifier that assigns the most frequent class to all observations [108]. Considering the correct predictions  $k = TP + TN$  as successes over the total of  $n$  predictions, a Binomial test was also performed:  $H_0 : \text{accuracy} = \text{NIR}$  against  $H_\alpha : \text{accuracy} \geq \text{NIR}$ . The associated  $p$ -value is:

$$p = \sum_k^n \binom{n}{k} \text{NIR}^k (1 - \text{NIR})^{n-k}$$

Under  $H_0 : p \sim B(n, p)$ . Under  $H_\alpha$ ,  $p$  tends to smaller values, in which case the model classifies better than "blindly" assigning the majority class. In other words, it represents the probability that the difference between accuracy and NIR is due to chance [111].

**2.4.3. Metabolite Network Analysis.** Network science presents a unifying framework for data and system representation, applicable to any domain [112]. A network, in an abstract sense, consists of nodes connected with links, also referred to as edges. In data science, a network whose nodes represent random features, whose joint probability distribution is defined by the ensemble of their edges is called *graphical model* [82]. A metabolomic covariance correlation network represents the ensemble of metabolites based on their covariance, showing nuances among the samples that non-graphical statistical methods on individual metabolites may fail to detect [113]. It may provide insights into correlated metabolites that do not belong in the same metabolic pathway [113].

A *Gaussian graphical model* (GGM) is an undirected graph that represents the conditional independence properties of the features [114]. The statistic employed by GGMs is the partial correlation which also adjusts for indirect correlation, i.e. two metabolites are correlated with a third one and are shown correlated with each other [115]. For instance, let  $\mathcal{V}$  a set of  $p$  vertices, representing random features  $Y_1, \dots, Y_p$  with joint probability distribution  $P \sim N_p(\mathbf{0}, \Sigma)$ , and  $\mathcal{E}$  set of edges, then  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is a GGM if for all pairs  $\{Y_i, Y_j\}$  with  $i \neq j$ :

$$\Sigma_{ij}^{-1} = (\Omega_{ij}) = 0 \iff Y_i \perp\!\!\!\perp Y_j \mid \{Y_k : k \neq i, j\} \iff (i, j) \notin \mathcal{E}.$$

In natural language, a zero value in the inverse covariance matrix (usually referred to as precision matrix  $\Omega$ ) implies that the respective random features are independent, given the rest of features, and they are not connected by an undirected edge  $((i, j) \notin \mathcal{E})$  [82].

For the metabolite network analysis, the package `rags2ridges` [82] was used to generate the feature covariance matrices of ApoE4 carriers and non-carriers with AD. Their precision matrices were *fused* to incorporate information from both groups, regularized with Ridge penalty and sparsified at a local FDR threshold  $q = 0.999$  using the `.fused` functions of `rags2ridges`. The optimal penalty was jointly estimated with 10-fold CV. The sparsified networks of each group, as well as their differential edges were plotted in GGMs using `Ugraph`. Lastly, communities of strongly correlated metabolites were identified using the community search algorithm by Girvan-Newman [116] as implemented in `rags2ridges`'s `Communities`. These metabolite ensembles may reflect conditionally dependent metabolite groups.

The network centrality measures (degree, betweenness, eigenvector centrality, number of positive and negative edges, mutual information and (partial) variance) were calculated using `GGMnetworkStats.fused`. Degree (of centrality) represents the sum of connected vertices [117]; each neighbouring node receives one "point". Eigenvector centrality, instead of awarding vertices just one point per neighbour, it gives each vertex a score proportional to the sum of the scores of its neighbours [117]. Betweenness reflects the extent to which a node lies on paths between other vertices [117]. Mutual information represents the amount of information we obtain for a node considering the rest of the nodes. The aforementioned measures aid in highlighting hub (central) metabolites with a regulatory potential. Wilcoxon Signed Rank tests were performed to test for differences between the network statistics and corrected for FDR at 0.05.

### 3. RESULTS

The ApoE genotype frequencies in AD and SCD are shown in Table 2. The most common genotype was  $\epsilon_3/\epsilon_3$ , followed by  $\epsilon_3/\epsilon_4$  and  $\epsilon_4/\epsilon_4$ . Cumulatively, the most abundant allele is  $\epsilon_3$  (64%), followed by  $\epsilon_4$  (30%) and  $\epsilon_2$  (6%). In the AD group, the allelic frequencies were 74.6% for  $\epsilon_3$ , followed by 15.4% for  $\epsilon_4$  and 10% for  $\epsilon_2$ . In the SCD group,  $\epsilon_3$ ,  $\epsilon_4$  and  $\epsilon_2$  exhibited relative frequencies of 53.5%, 44.5% and 2%, respectively. The increased  $\epsilon_4$  (Fig. 8) frequency in SCD patients might reflect compensatory mechanisms in place in this group or an increased risk for AD in the future.

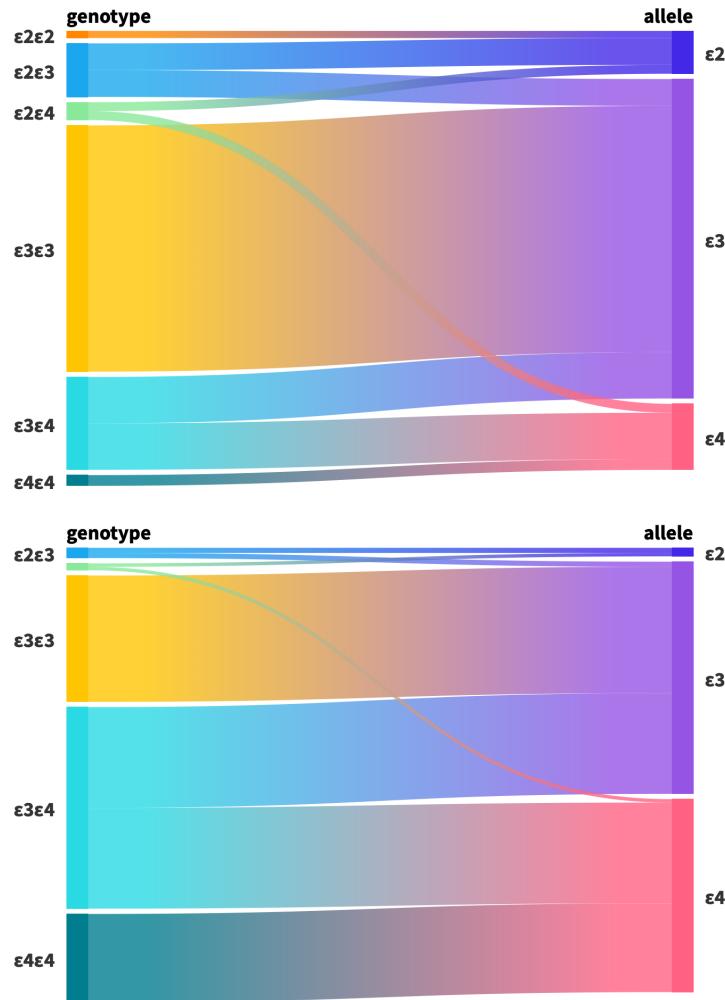
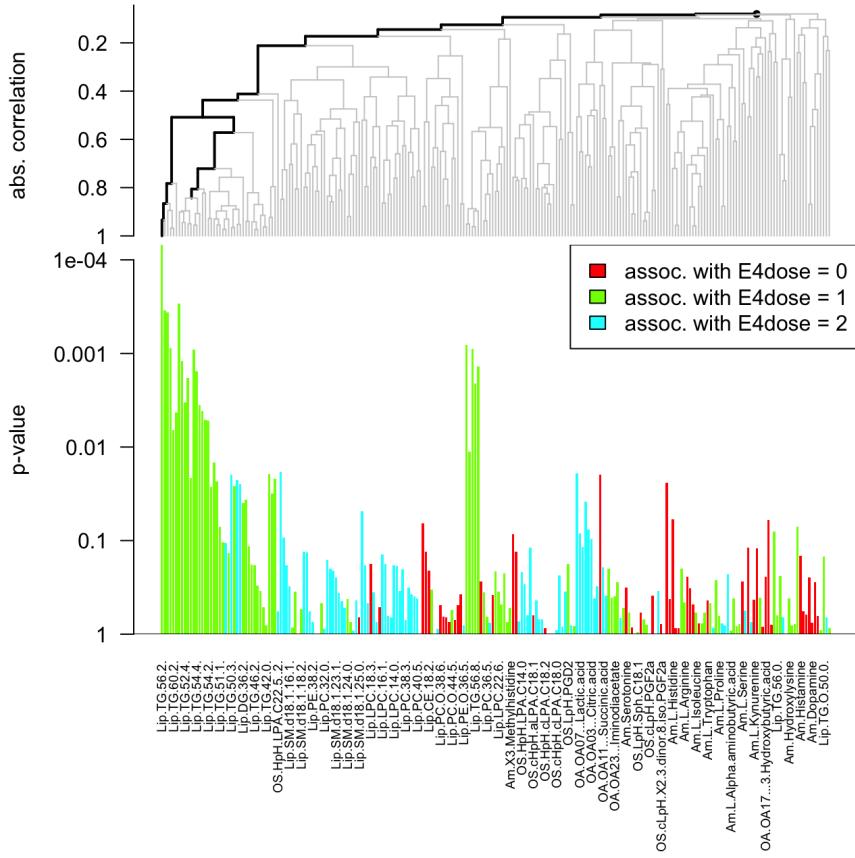


FIGURE 8. Sankey charts representing the ApoE allelic composition of the sample based on the genotype counts in AD (top) and SCD (bottom) (Table 2). Fig. 1 reflects the theoretical distribution of the three common ApoE alleles in  $\binom{3}{2}$  combinations (genotypes). This chart represents the relative ApoE allelic frequencies based on the 6 genotype counts from the data. ApoE3 is the most common allele, followed by ApoE4 and ApoE2 in both groups. A higher frequency of ApoE4 is observed in the SCD group compared to AD. This is attributed to the higher count of  $\epsilon_3/\epsilon_4$  and homozygous  $\epsilon_4$ .

### 3.1. ApoE4 dose effects on serum metabolite levels in AD.

3.1.1. *Global Test.* Testing for ApoE4 dose-effects on serum metabolites of AD patients, correcting for sex ( $H_0$ : ApoE4 dose has no effect on mean metabolite levels,  $H_a$ : it has an effect), revealed a significant global difference in metabolites ( $p = 0.017$ ). The most significantly altered metabolites are triglycerides and diglycerides (FDR-adjusted  $p$ -value  $<0.05$ ) See Table 3 and Fig. 9. In SCD individuals, the global test showed no significant differences among ApoE4 doses ( $p=0.544$ ). Generally, fewer metabolites were altered in this group and none of the associations remained significant after adjusting for FDR. Interestingly, the affected metabolites are different between AD and SCD.



**FIGURE 9.** Covariates plot showing the metabolites affected by the number of ApoE4 alleles in AD. The contribution of each such metabolite is itself a test, with a p-value against the alternative hypothesis (bottom). The plotted metabolites are ordered in a hierarchical clustering graph (top). The thick black line represents the metabolites and clusters of metabolites that are most clearly associated with ApoE4 dose, correcting for family-wise error rate at 0.05. This can be interpreted as that the multiple testing algorithm employed can confidently infer that at least one of the metabolites below the last significant branch is associated with the number of ApoE4 alleles, without being able to pinpoint with enough confidence which one(s).

TABLE 3. Metabolites affected by ApoE4 dose, correcting for sex as per globaltest [88]

	Class	Metabolite	Inheritance	Assoc. with	p-value	FDR
AD	Lipid	TG (56:2)	0.046	1 ApoE4	<0.001	0.016
	Lipid	TG (58:1)	0.117	1 ApoE4	<0.001	0.021
	Lipid	DG (36:3)	0.19	1 ApoE4	<0.001	0.021
	Lipid	TG (56:3)	0.244	1 ApoE4	<0.001	0.021
	Lipid	TG (52:3)	0.424	1 ApoE4	0.001	0.031
	Lipid	TG (54:5)	0.48	1 ApoE4	0.001	0.027
	Lipid	TG (58:2)	0.722	1 ApoE4	0.001	0.027
	Lipid	TG (54:4)	0.761	1 ApoE4	0.002	0.033
	Lipid	TG (58:9)	1	1 ApoE4	0.001	0.027
	Lipid	TG (56:7)	1	1 ApoE4	0.001	0.027
	Lipid	TG (58:8)	1	1 ApoE4	0.001	0.032
	Lipid	TG (54:6)	1	1 ApoE4	0.002	0.035
	Lipid	TG (56:8)	1	1 ApoE4	0.002	0.037
	Lipid	TG (52:4)	1	1 ApoE4	0.003	0.055
	Lipid	TG (54:3)	1	1 ApoE4	0.004	0.055
	Lipid	TG (56:6)	1	1 ApoE4	0.004	0.059
	Lipid	TG (56:1)	1	1 ApoE4	0.004	0.059
	Lipid	TG (52:2)	1	1 ApoE4	0.005	0.063
	Lipid	TG (54:2)	1	1 ApoE4	0.005	0.063
SCD	Lipid	TG (60:2)	1	1 ApoE4	0.007	0.077
	Lipid	TG (58:10)	1	1 ApoE4	0.011	0.124
	Lipid	TG (51:3)	1	1 ApoE4	0.015	0.154
	Lipid	SM (d18:1/18:1)	1	2 ApoE4	0.018	0.169
	Organic acid	2-ketoglutaric.acid	1	2 ApoE4	0.019	0.169
	Lipid	TG (52:0)	1	1 ApoE4	0.019	0.169
	Lipid	TG (50:3)	1	2 ApoE4	0.02	0.169
	Organic acid	Uracil	1	no ApoE4	0.02	0.169
	Lipid	TG (52:5)	1	1 ApoE4	0.021	0.174
	Lipid	TG (54:0)	1	1 ApoE4	0.022	0.174
Ox. Stress	Lipid	TG (50:2)	1	2 ApoE4	0.022	0.174
	Lipid	TG (51:2)	1	1 ApoE4	0.023	0.174
	Aminoacid	L-Glutamine	1	no ApoE4	0.024	0.174
	Lipid	TG (50:1)	1	2 ApoE4	0.025	0.174
	Lipid	TG (50:4)	1	1 ApoE4	0.026	0.176
	Lipid	TG (52:1)	1	1 ApoE4	0.027	0.176
	Lipid	TG (54:1)	1	1 ApoE4	0.031	0.203
	Lipid	TG (50:0)	1	1 ApoE4	0.037	0.229
	Organic acid	Malic.acid	1	2 ApoE4	0.038	0.234
	Lipid	DG (36:2)	1	1 ApoE4	0.039	0.235
CE	Ox. Stress	PAF (16:0)	1	2 ApoE4	0.049	0.283
	Lipid	LPC (20:5)	1	no ApoE4	0.012	0.828
	Lipid	SM (d18:1/23:0)	1	1 ApoE4	0.015	0.828
	Aminoacid	L-Tryptophan	1	no ApoE4	0.024	0.828
	Aminoacid	Putrescine	1	1 ApoE4	0.028	0.828
	Aminoacid	Glycine	1	1 ApoE4	0.035	0.828
	Lipid	CE (18:2)	1	1 ApoE4	0.038	0.828
	Aminoacid	SM (d18:1/22:0)	1	1 ApoE4	0.039	0.828
	Ox. Stress	LPA (20:5)	1	no ApoE4	0.040	0.828
	Lipid	PC (36:3)	1	1 ApoE4	0.043	0.828
DG	Organic acid	Succinic acid	1	1 ApoE4	0.043	0.828
	Amine	Citrulline	1	1 ApoE4	0.043	0.828

CE: Cholestryler ester, DG: Diglyceride, LPA: Lyso-sphingolipid LPC: Lysophosphatidylcholine, PAF: Platelet activating factor, SM: Sphingomyelin, TG: Triglyceride

**3.1.2. Nested Linear Models.** Several metabolites from all classes were altered by ApoE4 dose, both in AD and SCD. However, after controlling FDR, none of the effects remain significant at  $\alpha = 0.05$ . Among AD patients, metabolites showing trends of a positive effect were several triglycerides, diglycerides, putrescine, 2-ketoglutaric acid, lysophosphatidylcholin, platelet activating factor (16:0) and lyso-phosphatidic acid (18:0) (Table 4). Among individuals with SCD, lipid metabolites were not affected as much as in the AD group, with only two sphingomyelin species showing a difference. Aminoacids L-serine, tryptophan, glycine, tryptophan, L-homoserine, putrescine exhibited trends of effect in this group. L-Tryptophan is negatively associated with ApoE4 dose (at 1x and 2x ApoE4), while L-serine, glycine and L-homoserine are negatively associated only with ApoE4 homozygotes. (Table 4).

TABLE 4. ApoE4 dose effects on serum metabolites in AD and SCD: results from nested linear model comparison. Full model: clinical background variables and number of ApoE4 alleles, Nested model: clinical background variables only. Colour range green-red reflects trends in positive-negative correlations between ApoE4 dose and metabolites.

	Metabolite	P(>F)	FDR	No $\varepsilon_4$		1x $\varepsilon_4$		2x $\varepsilon_4$	
				Coef.	P(>t)*	Coef.	P(>t)*	Coef.	P(>t)*
AD	Lip.TG (52:3)	0.001	0.156	-1.4	0.818	2.6	0.004	3.7	0.001
	Lip.TG (52:4)	0.003	0.156	0.6	0.888	1.6	0.008	2.4	0.002
	DG (36:3)	0.002	0.156	0.0	0.631	0.0	0.012	0.0	0.001
	OS.HpH.PAF (16:0)	0.002	0.156	34.8	<0.001	2.6	0.017	4.6	0.001
	Lip.TG (52:2)	0.006	0.255	-5.1	0.469	2.1	0.03	3.7	0.002
	Lip.TG (54:5)	0.01	0.373	1.9	0.496	0.9	0.016	1.3	0.006
	Lip.TG (50:2)	0.012	0.398	-4.5	0.297	0.9	0.141	2.2	0.003
	Lip.TG (50:1)	0.018	0.45	-2.3	0.539	0.7	0.179	1.8	0.005
	Lip.TG (54:4)	0.02	0.45	3.5	0.318	1.2	0.015	1.4	0.02
	Lip.TG (54:6)	0.017	0.45	-0.4	0.795	0.5	0.027	0.7	0.009
	Am.Putrescine	0.029	0.515	0.0	<0.001	0.0	0.035	0.0	0.017
	OA.2.ketoglutaric.acid	0.026	0.515	0.0	0.366	0.0	0.953	0.0	0.014
	Lip.TG (50:3)	0.028	0.515	-1.2	0.668	0.6	0.127	1.3	0.008
	Lip.TG (52:5)	0.042	0.538	0.5	0.789	0.4	0.134	0.7	0.014
	Lip.TG (56:7)	0.042	0.538	-2.4	0.095	0.5	0.015	0.4	0.105
SCD	Lip.TG (56:8)	0.044	0.538	-1.4	0.055	0.2	0.014	0.2	0.151
	Lip.TG (58:9)	0.042	0.538	-0.4	0.068	0.1	0.013	0.0	0.41
	LPC (16:0)	0.033	0.538	4.2	0.028	0.3	0.237	0.9	0.009
	OS.HpH.LPA (18:0)	0.044	0.538	0.5	0.038	0.0	0.178	0.1	0.013
	Am.L-Serine	0.002	0.285	4.7	<0.001	0.2	0.115	-1.1	0.003
	Am.L-Tryptophan	0.002	0.285	3.7	0.004	-0.3	0.047	-1.4	0.002
	Am.Glycine	0.006	0.45	4.3	0.001	0.4	0.015	-0.9	0.052
SCD	Am.L-homoserine	0.047	0.931	0.0	0.001	0.0	0.723	0.0	0.014
	Am.Putrescine	0.045	0.931	0.0	0.969	0.0	0.013	0.0	0.927
	Lip.SM (d18:1/22:0)	0.043	0.931	3.3	0.002	0.3	0.015	0.3	0.448
	Lip.SM (d18:1/23:0)	0.029	0.931	1.2	0.01	0.1	0.012	0.2	0.277

\* Not corrected for multiple testing.

Am: Aminoacid, Lip: Lipid, DG: Diglyceride, LPA: Lyso-sphingolipid LPC: Lysophosphatidyl-choline, SM: Sphingomyelin, TG: Triglyceride, OA: Organic Acid, OS: Oxidative Stress compound, PAF: Platelet activating factor

**3.2. Classification of ApoE4 and AD status.** All classification models had a multi-class ROC AUC above 0.8 and accuracies significantly higher than the No-Information Rate ( $p < 2.2e^{-16}$ ) (Table 1) in predicting AD without ApoE4, AD with at least 1 ApoE4, SCD without ApoE4, SCD with at least 1 ApoE4. The worst-performing model was Multinomial Logistic Regression fitting the clinical data only, while the best-performing one was XGBoost fitting the clinical data and the 6 meta-metabolites (obtained from the Latent Factor Projection). Adding serum metabolite information (either the full 230-metabolite matrix or its 6-factor projection) seems to slightly increase the discriminatory power of the models. Notably, looking at the individual ROC curves (Fig. 10, 17 and Appendix B), as well as the confusion matrices (Fig. 12 and 13) all models were able to discriminate better among certain classes (AD+E4 vs. SCD+E4, AD+E4 vs. SCD, AD-E4 vs. SCD+E4 and AD-E4 vs. SCD-E4) compared to others (AD+E4 vs. AD-E4 and SCD+E4 vs. SCD-E4).

TABLE 5. Performance metrics of multi-class classification of ApoE4 and AD status, obtained from 10-fold CV repeated 100 times. The multi-class AUC is the average of the 6 one-vs-one classifications (see ROC curves (Fig. 10, 11 and Appendix B) for the binary AUCs)

Model	Fitting	AUC	Accuracy (95/% CI)	P(Acc. $\geq$ NIR)	NIR
MLR	Clinical features only	0.814	0.4807 (0.4744, 0.4869)	$< 2.2e^{-16}$	0.3522
	Clinical features + 230 metabolites	0.834	0.5753 (0.5692, 0.5815)	$< 2.2e^{-16}$	
		0.818	0.5313 (0.525, 0.5375)	$< 2.2e^{-16}$	
Decision Tree	Clinical features + 6 latent factors	0.818	0.5082 (0.502, 0.5145)	$< 2.2e^{-16}$	
	XGBoost	0.836	0.5944 (0.5882, 0.6005)	$< 2.2e^{-16}$	

AUC: multi-class Area Under the (ROC) Curve, MLR: Multinomial Logistic Regression, NIR: No-Information Rate

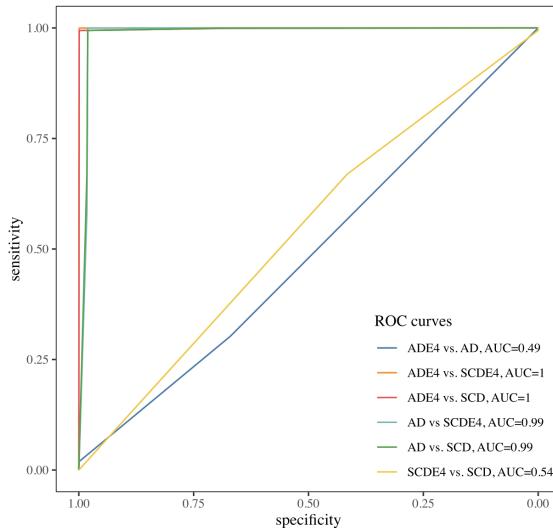


FIGURE 10. ROC curves of Multinomial Logistic Regression fitting the 230 metabolites on top of clinical background variables, obtained from repeated (100 times) 10-fold CV.

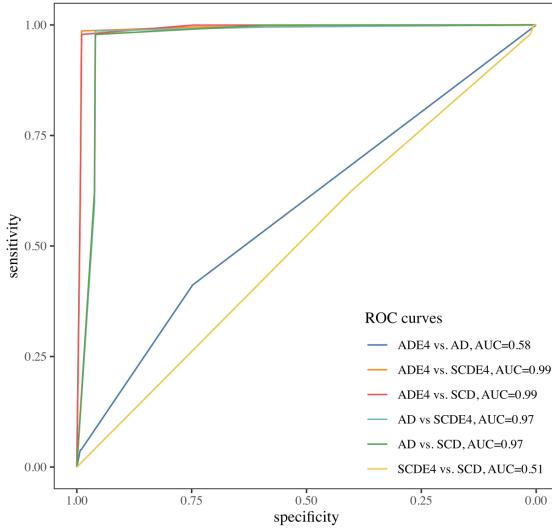


FIGURE 11. ROC curves of eXtreme gradient boosting fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV.



FIGURE 12. Confusion matrix of the true (Target) and predicted (Prediction) classes of Multinomial Logistic Regression fitting the 230 metabolites on top of clinical background variables, obtained from repeated (100 times) 10-fold CV. Each tile displays the *normalized* count (overall percentage) and the count of predicted classes in the middle. In the bottom of each tile the *column normalized* count reflects the proportion of the prediction over all predictions per true class. On the right of the tile, the row percentage shows the fraction of the prediction over the rest of true classes. The diagonal represents the correctly predicted classes (True Positive and True Negative).

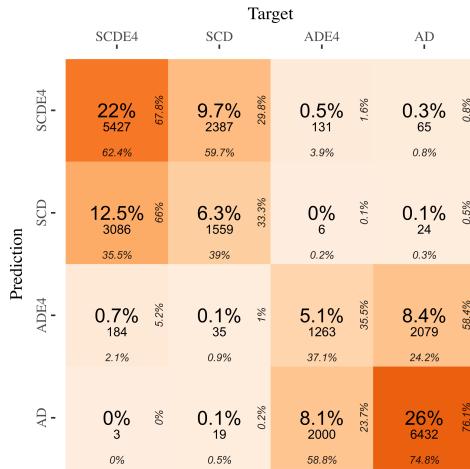


FIGURE 13. Confusion matrix of the true (Target) and predicted (Prediction) classes of XGBoost fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV.

TABLE 6. Features with the highest feature importance scores from a Multinomial Logistic Regression model fitting clinical background data and 230 metabolites to predicting ApoE4 and AD status, obtained via repeated (100 times) 10-fold CV.

Class	Feature	Overall (%)
Amine	Putrescine	100.00
Ox. stress	HpH.Spha.1.P.C18.0	97.43
Organic acid	Uracil	91.78
Lipid	TG (56:0)	88.86
Organic acid	3.Hydroxybutyric.acid	87.29
Clinical	Cholesterol medication	84.54
Amine	Sarcosine	83.54
Lipid	PE.O. (38:5)	81.60
Ox. stress	HpH.LPA.C20.5	72.43
Amine	L-Tryptophan	71.34
Lipid	SM (d18:1/20:1)	70.52
Amino acid	Glutathione	69.96
Amino acid	L-Serine	69.49
Amino acid	L-Histidine	69.41
Ox. stress	HpH.LPA.C18.0	68.81
Amino acid	Histamine	67.99
Amino acid	L-Threonine	67.85
Amino acid	L-Glutamine	67.70
Lipid	PC (38:3)	65.30
Organic acid	Pyruvic.acid	64.96

The most important metabolites in identifying the four ApoE4 and AD status phenotypes were putrescine, sphingosine 1-phosphate (18:0), uracil, triglyceride (56:0), 3- hydroxybutyric acid, sarcosine,

phosphoethanolamine (38:5), LPA (20:5), L-tryptophan, sphingomyelin (d18:1/20:1), glutathione, L-serine, L-histidine, LPA (18:0), histamine, L-threonine, L-glutamine, phosphatidylcholine (38:3), see Table 6.

**3.3. Metabolite Network Analysis.** The Gaussian graphical modelling between ApoE4 carriers and non-carriers with AD unveiled distinct correlations between the metabolites (Fig. 14). The ApoE4 carriers' network was more sparse, having less edges compared to the non-carriers'. Moreover, the metabolites could be clearly separated in communities based on their covariances in ApoE4 carriers, as shown in Fig. 15. The communities found contain similar compounds (phosphatidylcholines, triglycerides, sphingomyelines, prostaglandines) or molecules that share metabolic pathways (amines, amines-organic acids, aspartic-glutamic acid, oxidative stress compounds-lipids). The ApoE4 non-carrier's network appeared less cohesive, with only a few small aminic and organic acid communities.

The Wilcoxon Signed Rank tests revealed most network statistics to be larger in ApoE4 carriers compared to non-carriers (p-values <0.05) with AD (Table 7). In the ApoE4 positive group, top central metabolites were amines (dopamine and citrulline) (Table 8). ApoE4 non-carriers had 3-methoxytyrosine as top central metabolite, followed by the oxidative stress compound nitro fatty acid (C18:3).

TABLE 7. p-values of Wilcoxon Signed Rank test (alternative hypothesis: network statistics of ApoE4 carriers > ApoE4 non-carriers)

Statistic	p-value	FDR
Centrality	$8.459e^{-26}$	$1.973e^{-25}$
Betweennes	$4.110e^{-18}$	$5.754e^{-18}$
Eigenvector Centrality	$1.366e^{-06}$	$8.837e^{-06}$
Number of negative edges	$8.837e^{-06}$	$1.594e^{-06}$
Number of positive edges	$2.712e^{-23}$	$4.745e^{-23}$
Mutual Information	$4.087e^{-30}$	$1.603e^{-29}$
Variance	$4.582e^{-30}$	$1.603e^{-29}$

TABLE 8. Top central metabolites in ApoE4 carriers and non-carriers with AD.

ApoE4 carriers		ApoE4 non-carriers	
Metabolite	Degree	Metabolite	Degree
Am.Dopamine	6	Am.X3.Methoxytyrosine	5
Am.Citrulline	5	OS.LpH.NO2.OA	4
Am.X3.Methoxytyramine	4	Am.Cysteine	3
Am.ADMA	4	Am.Dopamine	3
Am.DL.3.aminoisobutyric.acid	4	Am.Glutathione	3
Am.Ethanolamine	4	Am.L.Aspartic.acid	3
Am.L.4.hydroxy.proline	4	Am.X3.Methylhistidine	2
Am.L.Kynurenone	4	Am.Hydroxylysine	2
Am.SDMA	4	Am.L.4.hydroxy.proline	2
OA.Glycolic.acid	4	Am.L.carnosine	2
OA.3.Hydroxybutyric.acid	4	Am.Methyldopa	2
OA.Aspartic.acid	4	OA.Citric.acid	2
OA.Glyceric.acid	4	OA.Pyruvic.acid	2
Lip.TG.56.8.	4	OA.Aspartic.acid	2
Lip.SM.d18.1.20.1.	4	OA.Iminodiacetate	2
OS.LpH.NO2.LA	4	OA.Uracil	2
OS.HpH.PAF.C16.0	4	Lip.TG.60.2.	2
OS.HpH.cLPA.C18.2	4	Lip.SM.d18.1.24.0.	2
Am.X3.Methoxytyrosine	3	OS.LpH.NO2.LA	2
Am.X3.Methylhistidine	3	OS.LpH.NO2.aLA	2
Am.gamma.aminobutyric.acid	3	OS.LpH.PGE2	2
Am.Glycine	3	OS.cLpH.PGA2	2
Am.L.2. amino adipic.acid	3	OS.cLpH.PGF2a	2
Am.L.Arginine	3	Am.X1.Methylhistidine	1
Am.L.Aspartic.acid	3	Am.Citrulline	1
Am.L.homoserine	3	Am.DL.3.aminoisobutyric.acid	1
Am.N6.N6.N6.Trimethyl.L.lysine	3	Am.Ethanolamine	1
Am.Sarcosine	3	Am.gamma.aminobutyric.acid	1
Am.Serotonin	3	Am.gamma.L. glutamyl.L.alanine	1
OA.Lactic.acid	3	Am.Glycylglycine	1
OA.Malic.acid	3	Am.Histamine	1
OA.2.ketoglutaric.acid	3	Am.L.2. amino adipic.acid	1
OA.Fumaric.acid	3	Am.L.Alanine	1
OA.Pyruvic.acid	3	Am.L.Alpha.aminobutyric.acid	1
OA.Pyroglutamic.acid	3	Am.L.Asparagine	1
OA.3.hydroxyisovaleric.acid	3	Am.L.Glutamic.acid	1
OA.Uracil	3	Am.L.homoserine	1
Lip.TG.42.0.	3	Am.L.Kynurenone	1

ADMA: Asymmetric Dimethylarginine, SD, Am: Aminoacid, Lip: Lipid, DG: Diglyceride, LPA: Lysophospholipid, SM: Sphingomyelin, TG: Triglyceride, OA: Organic Acid, OS: Oxidative Stress compound, PAF: Platelet Activating Factor, PGA2-PGE2: isoprostanes, NO2(a)La: nitro fatty acid, NO2.OA: nitro organic acid, SDMA: Symmetric Dimethylarginine.

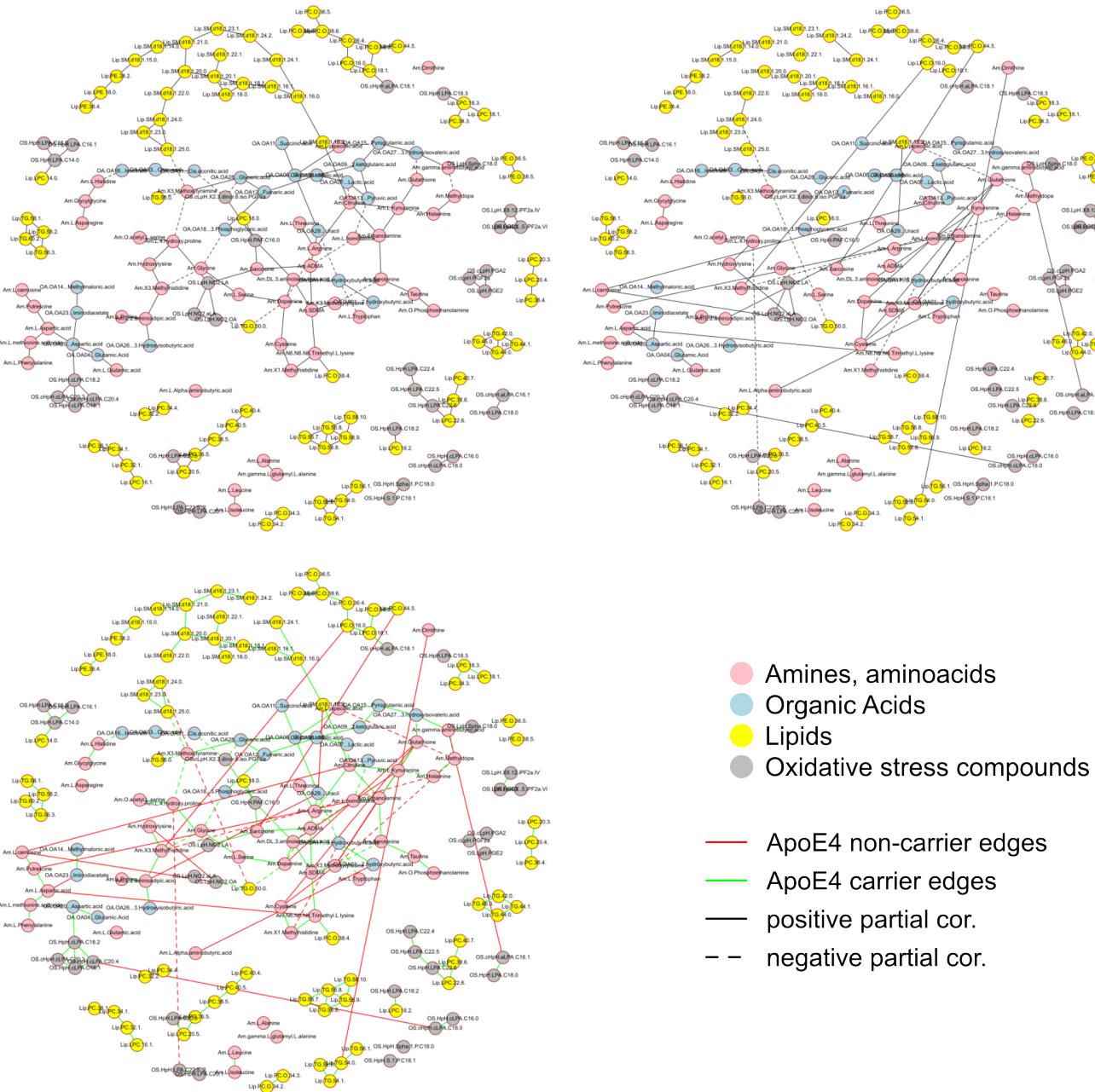


FIGURE 14. Metabolite precision network topologies among ApoE4 carriers (top-left), non-carriers (top-right) and differential edge network (bottom) in AD.

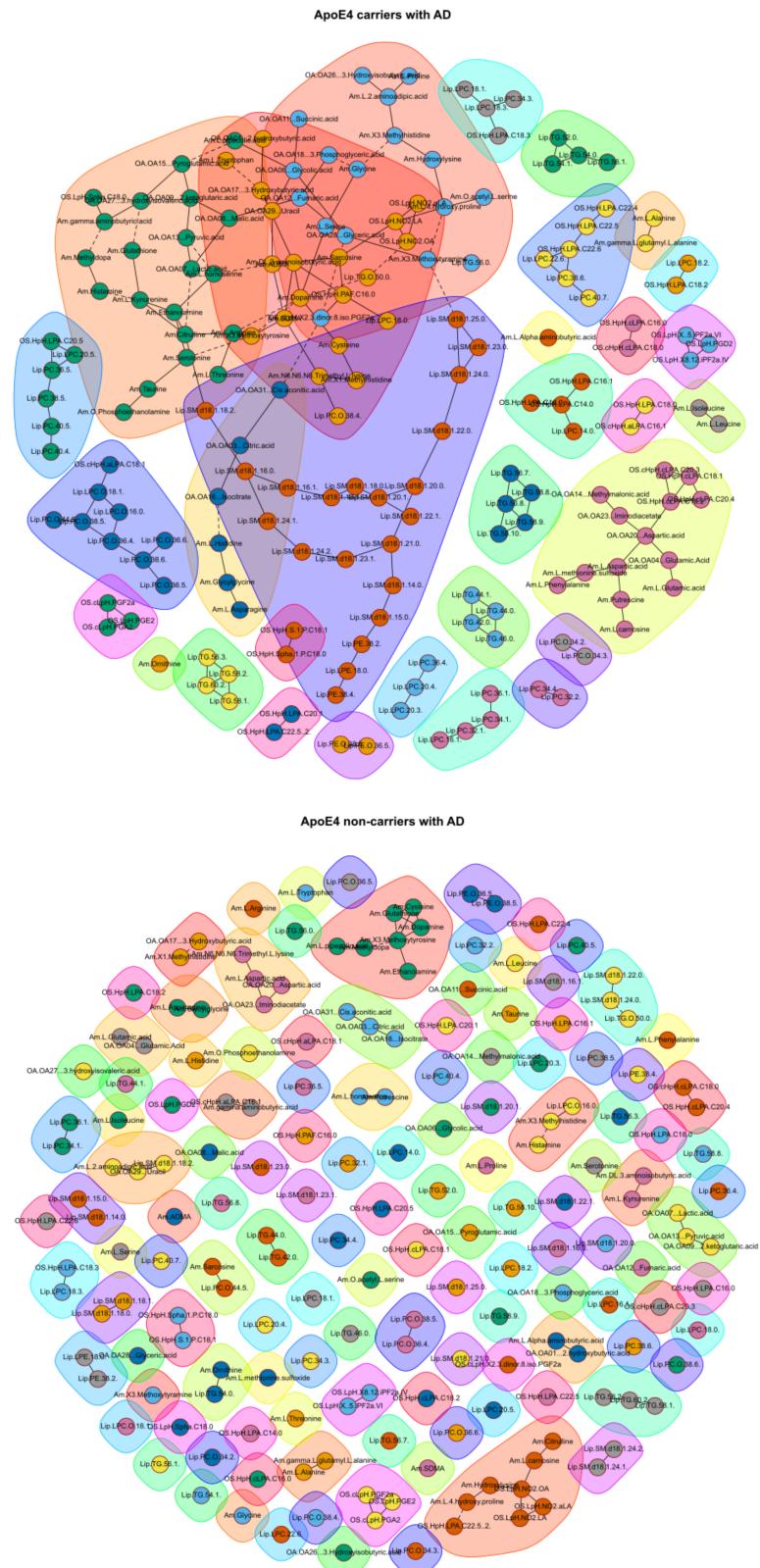


FIGURE 15. Communities of (densely correlated) metabolites in ApoE4 carriers (top) and ApoE4 non-carriers with AD (bottom) representing functional metabolic groups. The metabolome of ApoE4 carriers can be separated in communities, while those of ApoE4 non-carriers are less cohesive.

#### 4. DISCUSSION

This study had two objectives: to unveil mechanistic links between systematic metabolism and ApoE4 (number of alleles or status) in AD, and to propose a comprehensive omics data analysis methodology. To this end, three serum metabolic signatures were assessed among SCD and AD patients: individual metabolite levels, multi-class classification and network analysis. Individual metabolites were highlighted whose levels were shifted by ApoE4 dose. The multi-class classification revealed metabolites that, on top of clinical background variables, can be fitted to predict ApoE4 and AD status. Metabolite network analysis unravelled metabolic interdependencies among ApoE4 carriers and non-carriers in AD. The serum metabolites reported here might serve as a *proxy* of metabolic perturbations in the brain. The metabolites were measured in serum obtained from peripheral blood, an (almost) non-invasive alternative to spinal aspiration for CSF.

**4.1. Mechanistic links between ApoE4 dose and serum metabolome in AD.** The metabolites that emerged from the analysis were generally different among the approaches, and among AD-SCD. Nevertheless, certain metabolites appeared in all, or many of the metabolic signatures. The differential expression signature -as expected- revealed increased triglycerides and diglycerides, while the classification and network analysis -strikingly- presented a more pronounced aminic and organic acid signature.

A metabolite that appeared in all three metabolic signatures was putrescine: showed trends of positive ApoE4 dose effects in AD, was the most important feature in determining ApoE4 and AD status and central metabolite (5 dependent metabolites regardless of ApoE4 status) in AD. Putrescine appeared increased in a recent study [118], where it is described as a toxic by-product of activated urea cycle in A $\beta$ -reactive astrocytes of AD brains [118, 119].

Trends of decreased putrescine and four amino acids with increasing ApoE4 presence were observed in the SCD group: L-serine, L-tryptophan, glycine and L-homoserine. Tryptophan and serine proved also important in predicting ApoE4 and AD status. In the brain, tryptophan is catabolized to kynurene by the enzymes indoleamine 2,3-dioxygenase and tryptophan 2,3-dioxygenase, through the kynurene pathway [120]. L-tryptophan levels were positively correlated with no ApoE4 in this study, and negatively correlated with 1 or 2 ApoE4 alleles. This could be potentially explained by increased catabolism via the kynurene pathway. In this study, L-kynurene was found central in the serum metabolome of ApoE4 carriers with AD, branched with five metabolites. Recent reviews discussed kynurene pathway metabolites to be closely associated with AD pathogenesis [120, 121]. Increased kynurene pathway metabolites may reflect the extent of neuroinflammation in amyotrophic lateral sclerosis, frontotemporal dementia and early onset AD [122]. A whole-blood targeted metabolomics panel also revealed increased kynurene pathway metabolites [123]. Conversely, L-serine levels appear to be deficient in AD brains due to impaired glycolysis [124].

Ketoglutarate exhibited trends of decreased levels with increasing ApoE4 presence in the global test and the ANCOVA F-tests exclusively in the AD group. The metabolite network analysis revealed it is also a central metabolite, driving the variance of six other metabolites. In the brain, the  $\alpha$ -ketoglutarate dehydrogenase complex is perceived as a "hub of plasticity in neurodegeneration and regeneration", as it reflects the impaired glucose metabolism and key enzymes of the tricarboxylic acid cycle [125].

**4.1.1. Are there ApoE4 dose effects on serum metabolite levels in AD?** The global test for ApoE4 dose effects on serum metabolite levels revealed positive dose effects on lipids in AD, mainly triglycerides and diglycerides. This is consistent with evidence from Caucasian [126–128] and southern Chinese populations [129] that shows elevated triglycerides associated with ApoE4. The same test was not significant in the SCD group; mostly amines and amino acids showed trends of an effect. This could putatively reflect compensatory mechanisms still in place for the debilitating effects of ApoE4.

The nested linear model comparison did not reveal any significant ApoE4 dose effects, after correcting FDR. Even though the methods are fundamentally different, this can be partially explained considering the covariates in each approach. The global test was corrected for sex only, while the nested models fitted

several background clinical factors. That is, the metabolic variance explained by ApoE4 dose in the global test might be partially attributed to other clinical factors in the nested models, such as hypercholesterolemia. Notably, the metabolites showing trends of an ApoE4 dose effect were different between the AD and SCD group. Consistent with the global test, AD group presented mainly a lipid (tri- and diglycerides) signature, while the SCD group an amino one (serine, tryptophan, glycine, homoserine and putrescine).

**4.1.2. What is the (added) classification potential of serum metabolites in predicting ApoE4 and AD status?** Serum metabolic information, in the form of a semi-targeted metabolomic panel or its latent 6-factor projection seems to slightly increase the classification performance of the models. The added performance, however, is minor. As expected, XGBoost on clinical data and 6 meta-metabolites performed marginally better than penalized multinomial logistic regression on the clinical data and 230 metabolites, at the cost, however, of a significant loss of interpretability. Interestingly, all models discriminated better among certain classes (AD+E4 vs. SCD+E4, AD+E4 vs. SCD, AD-E4 vs. SCD+E4 and AD-E4 vs. SCD-E4) compared to others (AD+E4 vs. AD-E4 and SCD+E4 vs. SCD-E4). This can be putatively attributed to the class imbalance between AD+E4 ( $n = 34$ ) and AD ( $n = 86$ ), as well as SCD+E4 ( $n = 87$ ) and SCD ( $n = 40$ ). It might also show that the metabolic nuances among ApoE4 carriers and non-carriers are not as pronounced as among AD and SCD.

Each of the top four metabolites in delineating ApoE4 and AD status belonged to one of the major metabolite classes measured: amines (putrescine), oxidative stress compounds (sphingosine-1-phosphate (C18:0)), organic acids (uracil) and lipids (TG (56:0)). Perturbations in sphingolipid metabolism are observed in AD [130], even in the stage of MCI [131]. Den Hoedt *et al.* reported increased sphingosine-1-phosphate in CSF and plasma of homozygous ApoE4 carriers [131]. Another recent study highlights sphingolipids and especially sphingosine-1-phosphate as diagnostic markers for AD [132].

**4.1.3. How do the metabolite network topologies differ between ApoE4 carriers and non-carriers?** The GGMs revealed largely different metabolite network topologies between ApoE4 carriers and non-carriers. The metabolite network of ApoE4 carriers was less cohesive compared to non-carriers', in line with de Leeuw *et al.*'s findings. Further, the metabolite communities of the first include either molecules of the same biochemical subclass (e.g. phosphatidylcholines, sphingomyelins) or molecules sharing metabolic pathways (amines-organic acids, lipids-oxidative stress compounds) and might reflect functional groups. All the calculated metabolite network statistics were significantly larger in ApoE4 carriers compared to non-carriers. Amines are the top central metabolites in ApoE4 carriers, followed by organic acids. Dopamine was the most central metabolite in ApoE4 carriers. In this regard, evidence shows decreased dopaminergic neurotransmitters in AD [133, 134].

**4.2. Strengths and Limitations.** This analytical study was based on -and thus enabled by- uniquely detailed datasets in AD research. The  $n = 247$  subjects included in this study were screened for AD biomarkers in CSF, possible comorbidities, as well as several other risk factors. The clinical data alone were able to predict ApoE4 and AD status with an AUC of 81%. In this regard, another strength of this study is the comprehensive and efficient methodology it presents for -omics data analysis. The employed statistical methods here comprise a state-of-the-art analytical framework for metabolomics (or any -omics) data, to extract insights or complement diagnostic protocols. An innovation of this study is the use of a four-class target (outcome, response) in classification models, instead of a binary one (AD yes or no). To reach a high multi-class AUC, all the binary AUCs should be large. In this sense, the multi-class classification models predicting ApoE4 and AD status reached similar AUCs, either fitting the 230 metabolites (in multinomial logistic regression) or six meta-metabolites accounting for ~30% of their variance (in eXtreme gradient boosting).

AD is a complex and multifaceted disease with several proposed pathophysiological mechanisms. Hence, focusing on links between ApoE4 (status or dose) and serum metabolome in AD provides a limited view of the underlying pathologies at brain level. In this sense, gaining insights in ApoE genotype effects

via an untargeted metabolomics panel might prove more informative. Nevertheless, the relative genotype counts did not allow further analysis without grouping. Further, the metabolites were measured in blood -rather than CSF- which only reveals metabolic perturbations at systematic level. Another limitation is the fact that the metabolomics data were semi-target, thus several potentially important metabolites were not measured. Additionally, the significantly increased tri- and diglycerides in the global test might be attributed to hypertriglyceremia (that often coexists with hypercholesterolemia) and was not measured in this study. Moreover, it is important to note that the SCD group in the data is not a *control* group, but it represented subjects with varying degrees of subjective cognitive impairment, who might or might not develop AD or another neurodegenerative disease at later stages. Lastly, another limitation of the study is that the blood samples were obtained without prior fasting.

**4.3. Future directions.** The metabolites reported in this study (putrescine and sphingosine 1-phosphate) can be targeted in cohort studies to measure their levels in CSF and blood, throughout the AD continuum. In an effort to further elucidate molecular mechanisms implicated in AD development, further research should integrate other -omics data, over time. Such studies might provide additional stage-dependent AD biomarkers that can then be monitored in large-scale cohorts. Regarding the ApoE4 lipoprotein, it could be insightful to measure its levels in CSF and serum in different stages of AD, while considering other risk factors that modify the increased LOAD risk (ancestry, sex, or other inherited gene variants). Moreover, to allow further assessment of its effects, it is proposed to stratify per ApoE4 gene dose, status, or ApoE genotype during participant selection and sampling. Lastly, it could be insightful for disease predicting models to integrate as many of the reported AD-related biomarkers as possible. In this fashion, their relative diagnostic potential can be assessed.

**4.3.1. Statistical Analysis.** Multi-class classification models in AI-assisted differential diagnosis support should be preferred to binary, as they can predict a plurality of disease outcomes and better reflect reality. The R package *globaltest* offers a simple but robust approach to test for nuances in metabolite levels (or any biological high-dimensional data) attributed to a (multi-class) clinical feature -e.g. number of ApoE4 alleles. Even though *caret* is not the newest package for classification and regression pipelines in R, it is very simple, intuitive, well documented and maintained and is thus recommended for statistical learning research.

## 5. CONCLUSION

The serum metabolome is reported shifted by ApoE4 presence or dose. The multi-class classification of ApoE4 status and AD, as well as the metabolite network analysis provide additional metabolic signatures in the periphery. Tri- and diglycerides seem to be increased with increasing number of ApoE4 alleles in AD. The added classification potential of serum metabolome in identifying ApoE4 and AD status is small, but significant. Nevertheless, putrescine was the top metabolite in predicting ApoE4 and AD status, and thus might be worthwhile to assess in CSF at different stages of AD progression. Finally, the metabolite network analysis revealed distinct constellations of metabolites between ApoE4 carriers and non-carriers: metabolites of ApoE4 carriers could be grouped in communities of correlated metabolites, while among ApoE4 non-carriers not. Dopamine seems to be the most central metabolite in ApoE4 carriers with AD. Nevertheless, further investigation is needed to validate these findings, both in CSF and *post-mortem* brain tissues of AD patients, as well as in age-matched healthy controls having ApoE4 allelic presence, in service of a control group.

## REFERENCES

- Penke, B., Szűcs, M. & Bogár, F. New Pathways Identify Novel Drug Targets for the Prevention and Treatment of Alzheimer's Disease. *International Journal of Molecular Sciences* **24**, 5383. issn: 1422-0067. <https://www.mdpi.com/1422-0067/24/6/5383> (Mar. 2023).
- 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **19**, 1598–1695. issn: 1552-5279. <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.13016> (Apr. 2023).
- Heneka, M. T. et al. Neuroinflammation in Alzheimer's disease. *The Lancet. Neurology* **14**, 388–405. issn: 1474-4465. <https://pubmed.ncbi.nlm.nih.gov/25792098/> (Apr. 2015).
- Edwards, F. A. A Unifying Hypothesis for Alzheimer's Disease: From Plaques to Neurodegeneration. *Trends in Neurosciences* **42**, 310–322. issn: 1878108X. [https://www.cell.com/trends/neurosciences/abstract/S0166-2236\(19\)30027-X](https://www.cell.com/trends/neurosciences/abstract/S0166-2236(19)30027-X) (May 2019).
- Aaldijk, E. & Vermeiren, Y. The role of serotonin within the microbiota-gut-brain axis in the development of Alzheimer's disease: A narrative review. *Ageing Research Reviews* **75**, 101556. issn: 1568-1637. <https://www.sciencedirect.com/science/article/pii/S1568163721003032> (2022).
- Scheltens, P. et al. Alzheimer's disease. *Lancet (London, England)* **388**, 505–517. issn: 1474-547X. <https://pubmed.ncbi.nlm.nih.gov/26921134/> (July 2016).
- Beydoun, M. A. et al. Epidemiologic studies of modifiable factors associated with cognition and dementia: Systematic review and meta-analysis. *BMC Public Health* **14**, 1–33. issn: 14712458. <https://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-14-643> (June 2014).
- Van Cauwenbergh, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genetics in Medicine* **2016 18:5** **18**, 421–430. issn: 1530-0366. <https://www.nature.com/articles/gim2015117> (Aug. 2015).
- Hardy, J. Alzheimer's disease: The amyloid cascade hypothesis: An update and reappraisal. *Journal of Alzheimer's Disease* **9**, 151–153. issn: 1387-2877 (Jan. 2006).
- Kepp, K. P., Robakis, N. K., Høilund-Carlsen, P. F., Sensi, S. L. & Vissel, B. The amyloid cascade hypothesis: an updated critical review. <https://doi.org/10.1093/brain/awad159> (2023).
- Kurkinen, M. et al. The Amyloid Cascade Hypothesis in Alzheimer's Disease: Should We Change Our Thinking? *Biomolecules* **2023, Vol. 13, Page 453** **13**, 453. issn: 2218-273X. <https://www.mdpi.com/2218-273X/13/3/453> (Mar. 2023).
- 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. issn: 00368075 (Aug. 1993).
- Husain, M. A., Laurent, B. & Plourde, M. APOE and Alzheimer's Disease: From Lipid Transport to Physiopathology and Therapeutics. *Frontiers in Neuroscience* **15**, 630502. issn: 1662453X (Feb. 2021).
- Yang, L. G., March, Z. M., Stephenson, R. A. & Narayan, P. S. Apolipoprotein E in lipid metabolism and neurodegenerative disease. *Trends in Endocrinology & Metabolism* **34**, 430–445. issn: 1043-2760. [https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760\(23\)00092-9](https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760(23)00092-9) (Aug. 2023).
- Liu, C. & al. et, e. Apolipoprotein E and Alzheimer disease: risk, mechanisms, and therapy. *Nat. Rev. Neurosci.* **9**, 106–118 (2013).
- Strittmatter, W. & al. et, e. Apolipoprotein E: high-avidity binding to  $\beta$ -amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 1977–1981 (1993).
- Deming, Y. & al. et, e. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. *Acta Neuropathol.* **133**, 839–856 (2017).
- Bellon, M. & al. et, e. A quarter century of APOE and Alzheimer's disease: progress to date and the path forward. *Neuron* **101**, 820–838 (2019).
- Farrer, L. & al. et, e. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *JAMA* **278**, 1349–1356 (1997).
- Bookheimer, S. Y. & Burggren, A. C. APOE-4 genotype and neurophysiological vulnerability to Alzheimer's and cognitive aging. *Annual review of clinical psychology* **5**, 343–62. <https://www.annualreviews.org/doi/10.1146/annurev-clinpsy.032408.153625> (2009).
- Kim, J., Basak, J. M. & Holtzman, D. M. The Role of Apolipoprotein E in Alzheimer's Disease. *Neuron* **63**, 287–303. <https://doi.org/10.1016/j.neuron.2009.06.026> (2009).
- Serrano-Pozo, A. & Growdon, J. Is Alzheimer's Disease Risk Modifiable? *Journal of Alzheimer's disease : JAD* **67 3**, 795–819. <https://content.iospress.com/articles/journal-of-alzheimers-disease/jad181028> (2019).
- Arnold, M. et al. Sex and APOE e4 genotype modify the Alzheimer's disease serum metabolome. *Nature Communications* **2020 11:1** **11**, 1–12. issn: 2041-1723. <https://www.nature.com/articles/s41467-020-14959-w> (Mar. 2020).
- Yassine, H. N. & Finch, C. E. APOE Alleles and Diet in Brain Aging and Alzheimer's Disease. *Frontiers in Aging Neuroscience* **12**, 544681. issn: 16634365 (June 2020).
- Egert, S., Rimbach, G. & Huebbe, P. ApoE genotype: from geographic distribution to function and responsiveness to dietary factors. *Proceedings of the Nutrition Society* **71**, 410–424. issn: 1475-2719. <https://www.cambridge.org/core/journals>

- [proceedings - of - the - nutrition - society / article / apoe - genotype - from - geographic - distribution - to - function - and - responsiveness - to - dietary - factors / D9F35B3FFE46F48CAFE131E0E94CC62C](https://onlinelibrary.wiley.com/doi/10.1002/ajpa.21298) (Aug. 2012).
26. Eisenberg, D. T., Kuizawa, C. W. & Hayes, M. G. Worldwide allele frequencies of the human apolipoprotein E gene: Climate, local adaptations, and evolutionary history. *American Journal of Physical Anthropology* **143**, 100–111. issn: 1096-8642. <https://onlinelibrary.wiley.com/doi/10.1002/ajpa.21298> (Sept. 2010).
  27. Logue, M. W. *et al.* A Comprehensive Genetic Association Study of Alzheimer Disease in African Americans. *Archives of Neurology* **68**, 1569–1579. issn: 0003-9942. <https://jamanetwork.com/journals/jamaneurology/fullarticle/1108047> (Dec. 2011).
  28. Blue, E. E., Horimoto, A. R., Mukherjee, S., Wijsman, E. M. & Thornton, T. A. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. *Alzheimer's & Dementia* **15**, 1524–1532. issn: 1552-5279. <https://alz-journals.onlinelibrary.wiley.com/doi/10.1016/j.jalz.2019.07.016> (Dec. 2019).
  29. Suchy-Dicey, A., Howard, B., Longstreth, W. T., Reiman, E. M. & Buchwald, D. APOE genotype, hippocampus, and cognitive markers of Alzheimer's disease in American Indians: Data from the Strong Heart Study. *Alzheimer's & Dementia* **18**, 2518–2526. issn: 1552-5279. <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12573> (Dec. 2022).
  30. Rajabli, F. *et al.* Ancestral origin of ApoE e4 Alzheimer disease risk in Puerto Rican and African American populations. *PLOS Genetics* **14**, e1007791. issn: 1553-7404. <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007791> (Dec. 2018).
  31. Naslavsky, M. S. *et al.* Global and local ancestry modulate APOE association with Alzheimer's neuropathology and cognitive outcomes in an admixed sample. *Molecular Psychiatry* **2022** *27*:11 27, 4800–4808. issn: 1476-5578. <https://www.nature.com/articles/s41380-022-01729-x> (Sept. 2022).
  32. Chen, H.-K. & al. et. e. Apolipoprotein E4 domain interaction mediates detrimental effects on mitochondria and is a potential therapeutic target for Alzheimer disease. *J. Biol. Chem.* **286**, 5215–5221 (2011).
  33. Hunsberger, H. C., Pinky, P. D., Smith, W., Suppiramaniam, V. & Reed, M. N. The role of APOE4 in Alzheimer's disease: strategies for future therapeutic interventions. *Neuronal Signaling* **3**, 1–15. issn: 20596553. <https://dx.doi.org/10.1042/NS20180203> (June 2019).
  34. Trumble, B. C. *et al.* Apolipoprotein E4 is associated with improved cognitive function in Amazonian forager-horticulturalists with a high parasite burden. *FASEB Journal* **31**, 1508–1515. issn: 15306860 (Apr. 2017).
  35. Finch, C. E. & Sapolsky, R. M. The evolution of Alzheimer disease, the reproductive schedule, and apoE isoforms. *Neurobiology of Aging* **20**, 407–428. issn: 01974580 (1999).
  36. Mahley, R. Central nervous system lipoproteins: ApoE and regulation of cholesterol metabolism. *Arterioscler. Thromb. Vasc. Biol.* **36**, 1305–1315 (2016).
  37. Lanfranco, M. & al. et. e. Expression and secretion of apoE isoforms in astrocytes and microglia during inflammation. *Glia* **69**, 1478–1493 (2021).
  38. Castellano, J. M. *et al.* Human apoE isoforms differentially regulate brain amyloid- $\beta$  peptide clearance. *Science Translational Medicine* **3**. issn: 19466234 (June 2011).
  39. Cruchaga, C. *et al.* Cerebrospinal fluid APOE levels: An endophenotype for genetic studies for Alzheimer's disease. *Human Molecular Genetics* **21**, 4558–4571. issn: 09646906 (Oct. 2012).
  40. Bu, G. APOE targeting strategy in Alzheimer's disease: lessons learned from protective variants. *Molecular Neurodegeneration* **17**, 1–4. issn: 17501326. <https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-022-00556-6> (Dec. 2022).
  41. Hu, J. *et al.* Opposing effects of viral mediated brain expression of apolipoprotein E2 (apoE2) and apoE4 on apoE lipidation and  $\text{A}\beta$  metabolism in apoE4-targeted replacement mice. *Molecular Neurodegeneration* **10**, 1–11. issn: 17501326. <https://molecularneurodegeneration.biomedcentral.com/articles/10.1186/s13024-015-0001-3> (Mar. 2015).
  42. Flowers, S. A. & Rebeck, G. W. APOE in the normal brain. *Neurobiology of Disease* **136**, 104724. issn: 0969-9961 (Mar. 2020).
  43. Courtney, R. & Landreth, G. E. LXR Regulation of Brain Cholesterol: From Development to Disease. *Trends in Endocrinology and Metabolism* **27**, 404–414. issn: 18793061. [https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760\(16\)30004-2](https://www.cell.com/trends/endocrinology-metabolism/abstract/S1043-2760(16)30004-2) (June 2016).
  44. Heinsinger, N. M., Gachechiladze, M. A. & Rebeck, G. W. Apolipoprotein E Genotype Affects Size of ApoE Complexes in Cerebrospinal Fluid. *Journal of Neuropathology & Experimental Neurology* **75**, 918–924. issn: 0022-3069. <https://dx.doi.org/10.1093/jnen/nlw067> (Oct. 2016).
  45. Nguyen, D. *et al.* Molecular basis for the differences in lipid and lipoprotein binding properties of human apolipoproteins E3 and E4. *Biochemistry* **49**, 10881–10889. issn: 00062960. <https://pubs.acs.org/doi/abs/10.1021/bi1017655> (Dec. 2010).
  46. Hatters, D. M., Peters-Libeu, C. A. & Weisgraber, K. H. Apolipoprotein E structure: insights into function. *Trends in Biochemical Sciences* **31**, 445–454. issn: 09680004 (Aug. 2006).
  47. Hatters, D. M., Zhong, N., Rutenber, E. & Weisgraber, K. H. Amino-terminal Domain Stability Mediates Apolipoprotein E Aggregation into Neurotoxic Fibrils. *Journal of Molecular Biology* **361**, 932–944. issn: 00222836 (Sept. 2006).

48. Rawat, V. *et al.* ApoE4 Alters ABCA1 Membrane Trafficking in Astrocytes. *Journal of Neuroscience* **39**, 9611–9622. issn: 0270-6474. <https://www.jneurosci.org/content/39/48/9611.abstract> (Nov. 2019).
49. Zhao, N. *et al.* Apolipoprotein E4 Impairs Neuronal Insulin Signaling by Trapping Insulin Receptor in the Endosomes. *Neuron* **96**, 115–129. issn: 1097-4199. <https://pubmed.ncbi.nlm.nih.gov/28957663/> (Sept. 2017).
50. Svennerholm, L., Boström, K. & Jungbjer, B. Changes in weight and compositions of major membrane components of human brain during the span of adult human life of Swedes. *Acta Neuropathologica* **94**, 345–352. issn: 00016322. <https://link.springer.com/article/10.1007/s004010050717> (Oct. 1997).
51. Marottoli, F. M. *et al.* Peripheral inflammation, apolipoprotein E4, and amyloid- $\beta$  interact to induce cognitive and cerebrovascular dysfunction. *ASN Neuro* **9**. issn: 17590914. <https://journals.sagepub.com/doi/10.1177/1759091417719201> (Aug. 2017).
52. Teng, Z. *et al.* ApoE Influences the Blood-Brain Barrier Through the NF- $\kappa$ B/MMP-9 Pathway After Traumatic Brain Injury. *Scientific Reports* **2017** *7:1* 7, 1–8. issn: 2045-2322. <https://www.nature.com/articles/s41598-017-06932-3> (July 2017).
53. Kloske, C. M. & Wilcock, D. M. The Important Interface Between Apolipoprotein E and Neuroinflammation in Alzheimer's Disease. *Frontiers in Immunology* **11**. issn: 16643224 (Apr. 2020).
54. Huang, Y. W. A., Zhou, B., Wernig, M. & Südhof, T. C. ApoE2, ApoE3, and ApoE4 Differentially Stimulate APP Transcription and A $\beta$  Secretion. *Cell* **168**, 427–441. issn: 10974172 (Jan. 2017).
55. Tachibana, M. *et al.* Rescuing effects of RXR agonist bexarotene on aging-related synapse loss depend on neuronal LRP1. *Experimental Neurology* **277**, 1–9. issn: 0014-4886 (Mar. 2016).
56. Safieh, M., Korczyn, A. D. & Michaelson, D. M. ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Medicine* **2019** *17:1* 17, 1–17. issn: 1741-7015. <https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1299-4> (Mar. 2019).
57. Cao, J. *et al.* ApoE4-associated phospholipid dysregulation contributes to development of Tau hyper-phosphorylation after traumatic brain injury. *Scientific Reports* **7**. issn: 20452322 (Dec. 2017).
58. Shi, Y. *et al.* ApoE4 markedly exacerbates tau-mediated neurodegeneration in a mouse model of tauopathy. *Nature* **2017** *549*:7673 **549**, 523–527. issn: 1476-4687. <https://www.nature.com/articles/nature24016> (Sept. 2017).
59. Vasilevskaya, A. *et al.* Interaction of APOE4 alleles and PET tau imaging in former contact sport athletes. *NeuroImage: Clinical* **26**, 102212. issn: 2213-1582 (Jan. 2020).
60. Wang, C. & al. et. e. Gain of toxic apolipoprotein E4 effects in human iPSC-derived neurons is ameliorated by a small-molecule structure corrector. *Nat. Med.* **24**, 647–657 (2018).
61. Hoe, H. S., Freeman, J. & Rebeck, G. W. Apolipoprotein e decreases tau kinases and phospho-tau levels in primary neurons. *Molecular Neurodegeneration* **1**. issn: 17501326 (2006).
62. Barupal, D. K. *et al.* Sets of coregulated serum lipids are associated with Alzheimer's disease pathophysiology. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **11**, 619–627. issn: 2352-8729 (Dec. 2019).
63. Fernández-Calle, R. *et al.* APOE in the bullseye of neurodegenerative diseases: impact of the APOE genotype in Alzheimer's disease pathology and brain diseases. *Molecular Neurodegeneration* **2022** *17:1* 17, 1–47. issn: 1750-1326. <https://link.springer.com/article/10.1186/s13024-022-00566-4> (Sept. 2022).
64. Proitsi, P. *et al.* Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis. *Alzheimer's & Dementia* **13**, 140–151. issn: 1552-5260 (Feb. 2017).
65. Oeckl, P. *et al.* Glial Fibrillary Acidic Protein in Serum is Increased in Alzheimer's Disease and Correlates with Cognitive Impairment. *Journal of Alzheimer's disease : JAD* **67**, 481–488. issn: 1875-8908. <https://pubmed.ncbi.nlm.nih.gov/30594925/> (2019).
66. Bandaru, V. V. R. *et al.* ApoE4 disrupts sterol and sphingolipid metabolism in Alzheimer's but not normal brain. *Neurobiology of Aging* **30**, 591–599. issn: 0197-4580 (Apr. 2009).
67. Novotny, B. C. *et al.* Metabolomic and lipidomic signatures in autosomal dominant and late-onset Alzheimer's disease brains. *Alzheimer's & Dementia* **19**, 1785–1799. issn: 1552-5279. <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12800> (May 2023).
68. Area-Gomez, E. *et al.* APOE4 is Associated with Differential Regional Vulnerability to Bioenergetic Deficits in Aged APOE Mice. *Scientific Reports* **2020** *10:1* 10, 1–20. issn: 2045-2322. <https://www.nature.com/articles/s41598-020-61142-8> (Mar. 2020).
69. Zhao, N. *et al.* Alzheimer's Risk Factors Age, APOE Genotype, and Sex Drive Distinct Molecular Pathways. *Neuron* **106**, 727–742. issn: 10974199. [https://www.cell.com/neuron/abstract/S0896-6273\(20\)30185-9](https://www.cell.com/neuron/abstract/S0896-6273(20)30185-9) (June 2020).
70. Peña-bautista carmen, c. *et al.* Metabolomics study to identify plasma biomarkers in alzheimer disease: ApoE genotype effect. *Journal of Pharmaceutical and Biomedical Analysis* **180**, 113088. issn: 0731-7085 (Feb. 2020).
71. Farmer, B. & al. et. e. APOE4 lowers energy expenditure in females and impairs glucose oxidation by increasing flux through aerobic glycolysis. *Mol. Neurodegener.* **16**, 62 (2021).
72. De Leeuw, F. A. *et al.* Blood-based metabolic signatures in Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **8**, 196–207. issn: 2352-8729 (Jan. 2017).

73. Green, R. E. *et al.* Investigating associations between blood metabolites, later life brain imaging measures, and genetic risk for Alzheimer's disease. *Alzheimer's Research and Therapy* **15**, 1–13. issn: 17589193. <https://alzres.biomedcentral.com/articles/10.1186/s13195-023-01184-y%20http://creativecommons.org/publicdomain/zero/1.0/> (Dec. 2023).
74. Varma, V. R. *et al.* Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study. *PLOS Medicine* **15**, e1002482. issn: 1549-1676. <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002482> (Jan. 2018).
75. Sun, L. *et al.* Association between Human Blood Metabolome and the Risk of Alzheimer's Disease. *Annals of Neurology* **92**, 756–767. issn: 1531-8249. <https://onlinelibrary.wiley.com/doi/10.1002/ana.26464> (Nov. 2022).
76. Jia, L. *et al.* A metabolite panel that differentiates Alzheimer's disease from other dementia types. *Alzheimer's & Dementia* **18**, 1345–1356. issn: 1552-5279. <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12484> (July 2022).
77. Huo, Z. *et al.* Brain and blood metabolome for Alzheimer's dementia: findings from a targeted metabolomics analysis. *Neurobiology of Aging* **86**, 123–133. issn: 0197-4580 (Feb. 2020).
78. Weng, W. C., Huang, W. Y., Tang, H. Y., Cheng, M. L. & Chen, K. H. The Differences of Serum Metabolites Between Patients With Early-Stage Alzheimer's Disease and Mild Cognitive Impairment. *Frontiers in Neurology* **10**, 482026. issn: 16642295 (Nov. 2019).
79. Simpson, B. N. *et al.* Blood metabolite markers of cognitive performance and brain function in aging. *Journal of Cerebral Blood Flow and Metabolism* **36**, 1212–1223. issn: 15597016. <https://journals.sagepub.com/doi/full/10.1177/0271678X15611678> (July 2016).
80. Oka, T., Matsuzawa, Y., Tsuneyoshi, M. & Nakamura, Y. Multiomics analysis to explore blood metabolite biomarkers in an Alzheimer's Disease Neuroimaging Initiative cohort. <https://www.researchsquare.com/article/rs-2973576/v1> (June 2023).
81. Peeters, C. F. W. *et al.* Stable prediction with radiomics data. <https://arxiv.org/abs/1903.11696v1> (Mar. 2019).
82. Peeters, C. F., Bilgrau, A. E. & van Wieringen, W. N. rags2ridges: A One-Stop-ℓ2-Shop for Graphical Modeling of High-Dimensional Precision Matrices. *Journal of Statistical Software* **102**, 1–32. issn: 1548-7660. <https://www.jstatsoft.org/index.php/jss/article/view/v102i04> (May 2022).
83. Van Der Flier, W. M. & Scheltens, P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. *Journal of Alzheimer's disease : JAD* **62**, 1091–1111. issn: 1875-8908. <https://pubmed.ncbi.nlm.nih.gov/29562540/> (2018).
84. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **2016** *3*:1 3, 1–9. issn: 2052-4463. <https://www.nature.com/articles/sdata201618> (Mar. 2016).
85. Simon, R. M. *et al.* Design and Analysis of DNA Microarray Investigations Series: Statistics for Biology and Health, 199 (2004).
86. Goeman, J. J., Van de Geer, S., De Kort, F. & van Houwelingen, H. C. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics (Oxford, England)* **20**, 93–99. issn: 1367-4803. <https://pubmed.ncbi.nlm.nih.gov/14693814/> (Jan. 2004).
87. Goeman, J. J., Van De Geer, S. A. & Van Houwelingen, H. C. Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 477–493. issn: 1467-9868. <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9868.2006.00551.x> (June 2006).
88. Goeman, J., Oosting, J., Finos, L., Solari, A. & Edelmann, D. The Global Test and the globaltest R package (2023).
89. Ott, R. L. & Longnecker, M. T. *An introduction to statistical methods and data analysis* (Cengage Learning, 2015).
90. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289–300. issn: 00359246 (1995).
91. James, G., Witten, D., Hastie, T. & Tibshirani, R. An Introduction to Statistical Learning with Applications in R Second Edition (2023).
92. Drummond, C. in *Encyclopedia of Machine Learning* (eds Sammut, C. & Webb, G. I.) 171–171 (Springer US, Boston, MA, 2010). isbn: 978-0-387-30164-8. [https://doi.org/10.1007/978-0-387-30164-8\\_111](https://doi.org/10.1007/978-0-387-30164-8_111).
93. Elshawi, R., Al-Mallah, M. H. & Sakr, S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Medical Informatics and Decision Making* **19**, 1–32. issn: 14726947. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0874-0> (July 2019).
94. Geman, S., Bienenstock, E. & Doursat, R. Neural Networks and the Bias/Variance Dilemma. *Neural Computation* **4**, 1–58. issn: 0899-7667. <https://dx.doi.org/10.1162/neco.1992.4.1.1> (Jan. 1992).
95. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* Fourth. ISBN 0-387-95457-0. <https://www.stats.ox.ac.uk/pub/MASS4/> (Springer, New York, 2002).
96. Cessie, S. I. & Houwelingen, J. C. v. Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **41**, 191–201. issn: 1467-9876. <https://rss.onlinelibrary.wiley.com/doi/10.2307/2347628> (Mar. 1992).
97. Song, Y. Y. & Lu, Y. Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry* **27**, 130. issn: 10020829. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/> (Apr. 2015).

98. Saputra, R. *et al.* Detecting Alzheimer's disease by the decision tree methods based on particle swarm optimization in *Journal of Physics: Conference Series* **1641** (2020), 012025.
99. Dana, A.-D. & Alashqur, A. Using decision tree classification to assist in the prediction of Alzheimer's disease in 2014 6th international conference on computer science and information technology (CSIT) (2014), 122–126.
100. Kumar, A. & Singh, T. R. A new decision tree to solve the puzzle of Alzheimer's disease pathogenesis through standard diagnosis scoring system. *Interdisciplinary Sciences: Computational Life Sciences* **9**, 107–115 (2017).
101. Mofrad, R. B. *et al.* Decision tree supports the interpretation of CSF biomarkers in Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **11**, 1–9 (2019).
102. Therneau, T. M., Atkinson, B., Ripley, B., *et al.* rpart: Recursive partitioning. *R package version 3* (2010).
103. Friedman, J., Hastie, T. & Tibshirani, R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics* **28**, 337–407. <https://doi.org/10.1214/aos/1016218223> (2000).
104. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **29**, 1189–1232. issn: 00905364. <http://www.jstor.org/stable/2699986> (2001).
105. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* **13-17-August-2016**, 785–794. <https://arxiv.org/abs/1603.02754v3> (Mar. 2016).
106. Zhang, Y. *et al.* A multiclass extreme gradient boosting model for evaluation of transcriptomic biomarkers in Alzheimer's disease prediction. *Neuroscience Letters* **821**, 137609 (2024).
107. Dara, O. A. *et al.* Alzheimer's Disease Diagnosis Using Machine Learning: A Survey. *Applied Sciences* **13**. issn: 2076-3417. <https://www.mdpi.com/2076-3417/13/14/8298> (2023).
108. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1–26. issn: 1548-7660. <https://www.jstatsoft.org/index.php/jss/article/view/v028i05> (Nov. 2008).
109. Torgo, L. *Data Mining with R, learning with case studies, 2nd edition* <http://ltorgo.github.io/DMwR2> (Chapman and Hall/CRC, 2016).
110. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
111. Bicego, M. & Mensi, A. Null/No Information Rate (NIR): a statistical test to assess if a classification accuracy is significant for a given problem (June 2023).
112. Barabási, A.-L. *Network Science* <http://networksciencebook.com/> (Cambridge University Press, 2015).
113. Perez De Souza, L., Alseekh, S., Brotman, Y. & Fernie, A. R. Network-based strategies in metabolomics data analysis and interpretation: from molecular networking to biological interpretation. *Expert Review of Proteomics* **17**, 243–255. issn: 17448387 (Apr. 2020).
114. Koller, D. & Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*.
115. Amara, A. *et al.* Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation. *Frontiers in Molecular Biosciences* **9**, 841373. issn: 2296889X (Mar. 2022).
116. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113. <https://link.aps.org/doi/10.1103/PhysRevE.69.026113> (2 Feb. 2004).
117. Newman, M. *Networks: An Introduction* ISBN: 9780199206650. <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001> (Oxford University Press, Mar. 2010).
118. Ju, Y. H. *et al.* Astrocytic urea cycle detoxifies A $\beta$ -derived ammonia while impairing memory in Alzheimer's disease. *bioRxiv*. <https://doi.org/10.1101/2021.10.15.464517> (2021).
119. Wong, W. Pathogenic putrescine. *Science signaling* **15** **751**, eade8161. <https://www.science.org/doi/10.1126/scisignal.ade8161> (2022).
120. Liang, Y. *et al.* Kynurenone Pathway Metabolites as Biomarkers in Alzheimer's Disease. *Disease Markers* **2022**. <https://doi.org/10.1155/2022%2F9484217> (2022).
121. Sharma, V., Singh, T. G., Prabhakar, N. K. & Mannan, A. V. Kynurenone Metabolism and Alzheimer's Disease: The Potential Targets and Approaches. *Neurochemical Research* **47**, 1459–1476. <https://doi.org/10.1007/s11064-022-03546-8> (2022).
122. Heylen, A. *et al.* Brain Kynurenone Pathway Metabolite Levels May Reflect Extent of Neuroinflammation in ALS, FTD and Early Onset AD. *Pharmaceuticals* **16**. issn: 1424-8247. <https://www.mdpi.com/1424-8247/16/4/615> (2023).
123. Teruya, T., Chen, Y.-J., Kondoh, H., Fukuiji, Y. & Yanagida, M. Whole-blood metabolomics of dementia patients reveal classes of disease-linked metabolites. *Proceedings of the National Academy of Sciences of the United States of America* **118**. <https://doi.org/10.1073/pnas.2022857118> (2021).
124. Douce, J. L. *et al.* Impairment of Glycolysis-Derived l-Serine Production in Astrocytes Contributes to Cognitive Deficits in Alzheimer's Disease. *Cell metabolism* **31** **3**, 503–517.e8. <https://doi.org/10.1016/j.cmet.2020.02.004> (2020).
125. Hansen, G. E. & Gibson, G. E. The  $\alpha$ -Ketoglutarate Dehydrogenase Complex as a Hub of Plasticity in Neurodegeneration and Regeneration. *International Journal of Molecular Sciences* **23**. <https://www.mdpi.com/1422-0067/23/20/12403> (2022).

126. Maxwell, T. J. *et al.* APOE Modulates the Correlation Between Triglycerides, Cholesterol, and CHD Through Pleiotropy, and Gene-by-Gene Interactions. *Genetics* **195**, 1397–1405. <http://www.genetics.org/content/genetics/early/2013/09/30/genetics.113.157719.full.pdf> (2013).
127. Carvalho-Wells, A. L., Jackson, K. G., Lockyer, S., Lovegrove, J. A. & Minihane, A. M. APOE genotype influences triglyceride and C-reactive protein responses to altered dietary fat intake in UK adults. *The American Journal of Clinical Nutrition* **96**, 1447–1453. ISSN: 0002-9165. <https://www.sciencedirect.com/science/article/pii/S0002916523029349> (2012).
128. Bernath, M. *et al.* Serum triglycerides in Alzheimer disease. *Neurology* **94**, e2088–e2098. <https://doi.org/10.1212/WNL.0000000000009436> (2020).
129. Gan, C., Zhang, Y., Liang, F., Guo, X. & Zhong, Z. Effects of APOE gene  $\epsilon$ 4 allele on serum lipid profiles and risk of cardiovascular disease and tumorigenesis in southern Chinese population. *World Journal of Surgical Oncology* **20**, 1–12. ISSN: 14777819. <https://wjso.biomedcentral.com/articles/10.1186/s12957-022-02748-2> (Dec. 2022).
130. Mielke, M. M. & Lyketsos, C. G. Alterations of the sphingolipid pathway in Alzheimer's disease: new biomarkers and treatment targets? *Neuromolecular medicine* **12**, 331–340 (2010).
131. Den Hoedt, S. *et al.* Sphingolipids in Cerebrospinal Fluid and Plasma Lipoproteins of APOE4 Homozygotes and Non-APOE4 Carriers with Mild Cognitive Impairment versus Subjective Cognitive Decline. *Journal of Alzheimer's Disease Reports*, 1–16 (2023).
132. D'Angiolini, S., Chiricosta, L. & Mazzon, E. Sphingolipid metabolism as a new predictive target correlated with aging and ad: a transcriptomic analysis. *Medicina* **58**, 493 (2022).
133. Shaikh, A., Ahmad, F., Teoh, S. L., Kumar, J. & Yahaya, M. F. Targeting dopamine transporter to ameliorate cognitive deficits in Alzheimer's disease. *Frontiers in Cellular Neuroscience* **17**, 1292858 (2023).
134. Pan, X. *et al.* Dopamine and dopamine receptors in Alzheimer's disease: A systematic review and network meta-analysis. *Frontiers in aging neuroscience* **11**, 175 (2019).

## APPENDIX A. CODE AND RESULTS

An HTML document with the code and the output of the analysis can be found at [gmiliarakis.github.io](https://gmiliarakis.github.io).

## APPENDIX B. MULTICLASS ROC CURVES

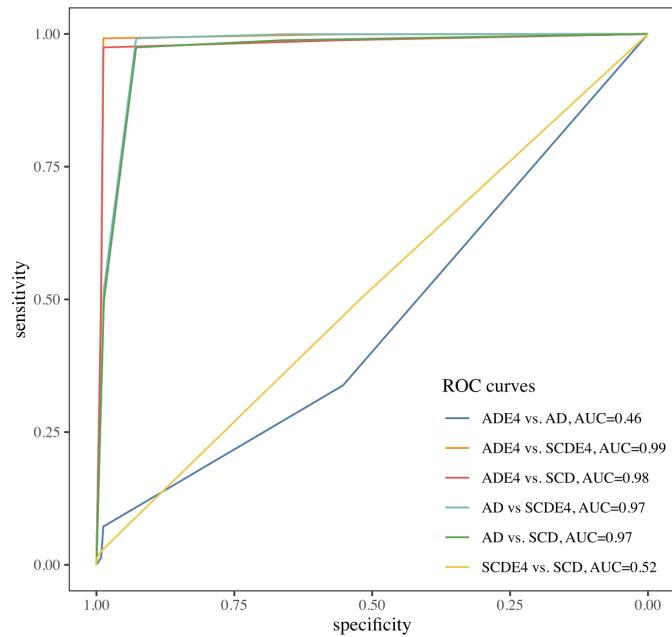


FIGURE 16. ROC curves of benchmark model: Multinomial Logistic Regression fitting the clinical background features only, obtained from repeated (100 times) 10-fold CV.

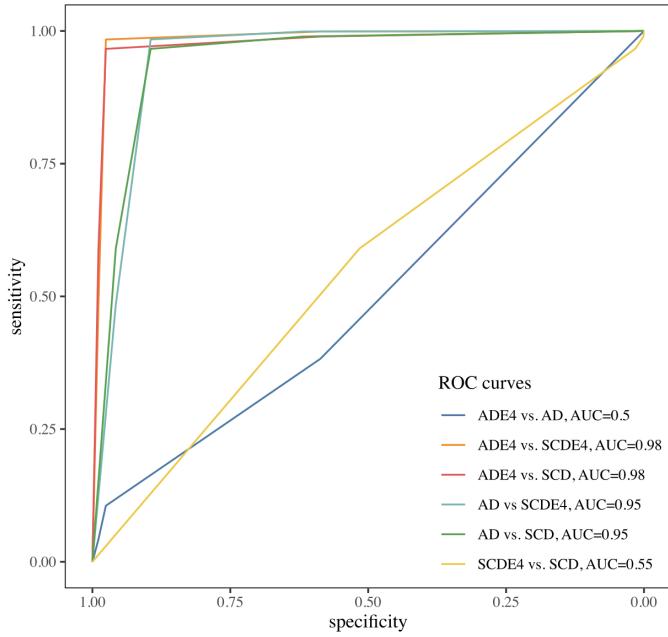


FIGURE 17. ROC curves of multinomial logistic regression fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV.

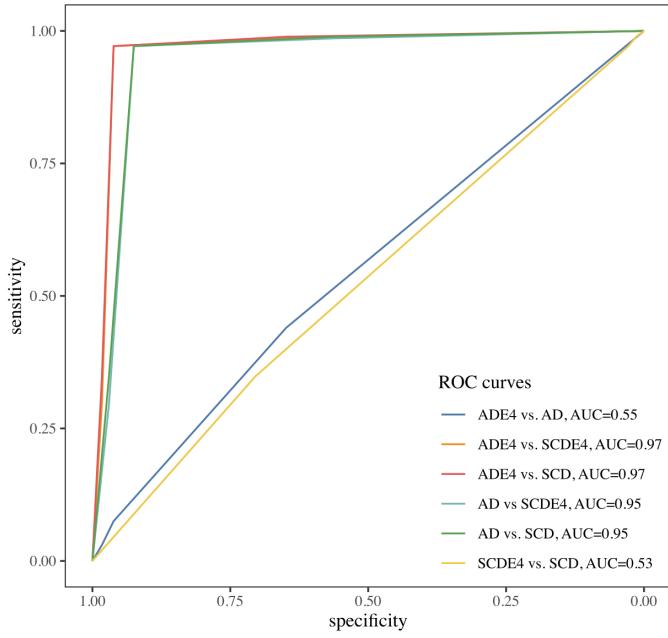


FIGURE 18. ROC curves of Decision Tree fitting the 6 ML-estimated latent factors on top of the clinical background variables, obtained from repeated (100 times) 10-fold CV.

## APPENDIX C. R SESSION INFORMATION

```
R version 4.3.2 (2023-10-31)
Platform: aarch64-apple-darwin20 (64-bit)
Running under: macOS Sonoma 14.2.1

Matrix products: default
BLAS:     .../vecLib.framework/Versions/A/libBLAS.dylib
LAPACK: LAPACK version 3.11.0

locale:
[1] en_US.UTF-8

time zone: Europe/Amsterdam
tzcode source: internal

attached base packages:
[1] stats      graphics   grDevices datasets  utils      methods    base

other attached packages:
[1] rags2ridges_2.2.7      nnet_7.3-19          DMwR2_0.0.2
[4] dplyr_1.1.4            caret_6.0-94         lattice_0.21-9
[7] globaltest_5.56.0      survival_3.5-7       heatmaply_1.5.0
[10] xgboost_1.7.6.1       pROC_1.18.5          FMradio_1.1.1
[13] future_1.33.0          ggthemes_5.0.0        corpcor_1.6.10
[16] viridisLite_0.4.2      plotly_4.10.3         ggplot2_3.4.4
[19] rpart_4.1.21           furrr_0.3.1          viridis_0.6.4
[22] e1071_1.7-14          cvms_1.6.0

loaded via a namespace (and not attached):
[1] splines_4.3.2          bitops_1.0-7          tibble_3.2.1
[4] XML_3.99-0.16          lifecycle_1.0.4       globals_0.16.2
[7] magrittr_2.0.3          Hmisc_5.1-1           rmarkdown_2.25
[10] lubridate_1.9.3         zlibbioc_1.48.0       sfsmisc_1.1-16
[13] RCurl_1.98-1.13        ipred_0.9-14          lava_1.7.3
[16] S4Vectors_0.40.2       listenv_0.9.0          gRbase_2.0.1
[19] codetools_0.2-19        tidyselect_1.2.0       TSP_1.2-4
[22] jsonlite_1.8.8         Formula_1.2-5         iterators_1.0.14
[25] snowfall_1.84-6.3      Rcpp_1.0.11           glue_1.6.2
[28] tufte_0.13             TTR_0.24.4            GenomeInfoDb_1.38.2
[31] fastmap_1.1.1          fansi_1.0.6           digest_0.6.33
[34] RSQLite_2.3.4           utf8_1.2.4            tidyR_1.3.0
[37] recipes_1.0.9          class_7.3-22          httr_1.4.7
[40] gtable_0.3.4           timeDate_4032.109    blob_1.2.4
[43] RBGL_1.78.0            GSEABase_1.64.0       scales_1.3.0
[46] knitr_1.45              rstudioapi_0.15.0     tzdb_0.4.0
[49] curl_5.2.0              proxy_0.4-27          cachem_1.0.8
[52] foreign_0.8-85          AnnotationDbi_1.64.1  pillar_1.9.0
[55] VGAM_1.1-9              xtable_1.8-4           cluster_2.1.4
```

```
[58] cli_3.6.2           compiler_4.3.2      rlang_1.1.2
[61] plyr_1.8.9          stringi_1.8.3      assertthat_0.2.1
[64] Matrix_1.6-1.1     hms_1.1.3          bit64_4.0.5
[67] quantmod_0.4.25    bit_4.0.5          hardhat_1.3.0
[70] graph_1.80.0       xts_0.13.1         MASS_7.3-60
[73] dendextend_1.17.1  backports_1.4.1   yaml_2.3.8
[76] DBI_1.1.3          RColorBrewer_1.1-3 expm_0.999-8
[79] purrr_1.0.2         BiocGenerics_0.48.1 seriation_1.5.4
[82] IRanges_2.36.0      RSpectra_0.16-1   GenomeInfoDbData_1.2.11
[85] annotate_1.80.0     parallelly_1.36.0 stats4_4.3.2
[88] foreach_1.5.2       tools_4.3.2        snow_0.4-4
[91] prodlim_2023.08.28 gridExtra_2.3    xfun_0.41
[94] ca_0.71.1          withr_2.5.2        BiocManager_1.30.22
[97] timechange_0.2.0    R6_2.5.1          colorspace_2.1-0
[100] generics_0.1.3     renv_1.0.3         data.table_1.14.10
[103] htmlwidgets_1.6.4   ModelMetrics_1.2.2.2 pkgconfig_2.0.3
[106] registry_0.5-1     XVector_0.42.0   htmltools_0.5.7
[109] Biobase_2.62.0     png_0.1-8         gower_1.0.1
[112] reshape2_1.4.4      checkmate_2.3.1  nlme_3.1-163
[115] zoo_1.8-12         stringr_1.5.1    parallel_4.3.2
[118] grid_4.3.2          reshape_0.8.9    vctrs_0.6.5
[121] htmlTable_2.4.2    evaluate_0.23   readr_2.1.4
[124] crayon_1.5.2       future.apply_1.11.0 fdrtool_1.2.17
[127] munsell_0.5.0       Biostrings_2.70.1 lazyeval_0.2.2
[130] KEGGREST_1.42.0    igraph_1.6.0     memoise_2.0.1
[133] base64enc_0.1-3    webshot_0.5.5
```