

Mechanistic links between ApoE genotype and serum metabolites in Alzheimer's Disease

a Data Science approach

George Miliarakis

MSc Research Proposal
Wageningen University & Research
Wageningen, November 2023

Supervisor

C.F.W. Peeters

Mathematical & Statistical Methods (Biometris)
Wageningen University & Research

Supervisor

Yannick Vermeiren

Nutrition, Brain and Cognitive Ageing
Wageningen University & Research

CONTENTS

1. INTRODUCTION

1.1. Background. Alzheimer's Disease (AD) is a complex, progressive neurodegenerative disorder and the most prevalent form of dementia [1]. One person is diagnosed with AD every 3 seconds, with a two-fold increase in incidence every 10 years, i.e., 10% at the age of 65, 20% at 75, and 40% at 85 [2]. The impact AD has on patients, their caregivers, and healthcare systems is detrimental. In fact, it has an estimated annual global societal cost that exceeds the GDP of The Netherlands [3]. To date, there is no successful disease or outcome modifying therapy for AD.

The main risk factor for AD is age, while several genetic and lifestyle risk factors, as well as biochemical pathways contribute to its development [1]. AD occurs in various histopathological phenotypes and presents a broad spectrum of clinical signs and symptoms [4], [5]. The Alzheimer's disease continuum starts with subjective cognitive decline (SCD), followed by mild cognitive impairment (MCI) [6], 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 [7].

The most prevalent phenotype of AD is sporadic or late-onset AD (SAD, LOAD; 95% of cases), typically appearing after 65 years of ages [8]. A rarer phenotype is early-onset familial AD (FAD), usually starting at ages 30–65 and passed in an autosomal dominant fashion [9]. Even though FAD mutations explain only a small percentage of AD cases, they have a great impact on AD research given their appealing genotype-phenotype links.

The amyloid cascade hypothesis has been the most predominant in explaining the pathogenesis of FAD. In historical terms, its impact was profound, as it helped distinguish and identify AD as a single disease that may be studied for treatment [10]. 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) [5]. Nevertheless, it does not necessarily cover all the AD cases; clinical trials of $A\beta$ anti-bodies as treatment prove the amyloid cascade hypothesis insufficient [11], [12]. Evidence suggests that oxidative stress, metabolic abnormalities, atherosclerosis, cardiovascular effects, imbalances of intra-neuron calcium and other metal ions contribute to the development of AD [11]. Kepp *et al.* in their recent review propose a more complex and holistic view of AD pathology, by integrating (epi-)genetic, environmental, vascular, neuro-inflammatory and metabolic factors in predictive models [11]. The present study focuses on integrating and linking genetic and metabolic nuances in AD.

Metabolism entails the repertoire of chemical reactions that keep living organisms alive. Metabolites—especially lipid [13]–[15]—, perceived as functional intermediates of AD development, are rigorously studied for bio-markers or targets for treatment [16].

The apo-lipoprotein- ϵ (ApoE) is a brain-specific fat-binding protein that comprises several types of lipoproteins found in blood, interstitial fluid (ISF) of brain parenchyma and CSF, i.e., chylomicra, IDL and VLDL [17]. Humans have three versions of the gene: ϵ_2 which is protective, ϵ_3 that is neutral and ϵ_4 that is predisposing of AD [17]. The ApoE polymorphism is found in 6 different genotypes [Fig. ??], and affects serum levels of several lipids [18].

The $APO\epsilon_4$ form has the highest lipid-binding affinity and leads to poor ApoE recycling, $A\beta$ clearance, glucose uptake and mitochondrial deregulation in the brain [17]. Moreover, ApoE4 is associated with tau misfolding and increased pro-inflammatory response (e.g. via microglia and astrocyte activation) which compromise the blood-brain barrier (BBB), leading to neurodegeneration and cognitive decline [17], [19]. Notably, sex (60% females) and ApoE4 allelic composition (50% has at least one ϵ_4 allele) are the strongest genetic risk factors for SAD [17], [20]. 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 [20].

Perturbed serum metabolites associated with AD are amines, aminoacids [21], [22], cholesteryl esters [15], sphingolipids [13], [16], [22]–[24], fatty acids [14], [21], glycerophospholipids [23], [25]–[27], phosphatidylcholines [28] and lipid peroxidation compounds [14]. These molecules are usually identified

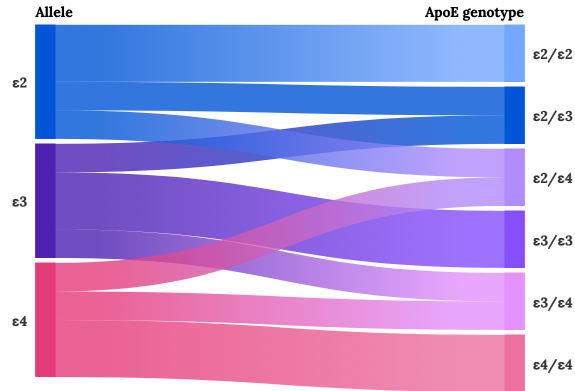


FIGURE 1. Sankey chart showing the allele distribution among the 6 ApoE genotypes.

via high-throughput metabolomic pipelines (coupled with Mass Spectrometry (MS) detectors) that trace all compounds in a sample and result in high-dimensional data [29]. The latter often require advanced statistical methods e.g. projection to latent structures [27], [30] or graphical models [31] in order to extract putatively meaningful information. With such techniques, Leeuw *et al.* discovered distinct serum metabolic signatures among AD patients-controls and those carrying at least one ApoE ϵ_4 allele [21], as they appear in Fig. ?? . The different metabolic profiles, however, among ApoE4 non-carriers remain obscure. A data science approach can potentially reveal differentially expressed metabolites between the 6 ApoE genotypes, thus unveiling distinct pathways of metabolic deregulation in AD.

1.2. Research Questions. ApoE ϵ_4 carriers –particularly females– experience metabolic disturbances and are at increased risk of SAD. The mechanistic links, however, between ApoE genotype, metabolism and AD development are not entirely known [14], more so among ϵ_4 non-carriers. Linking specific serum metabolites to distinct ApoE genotypes can reveal metabolic disturbances leading to AD, especially in absence of the ApoE ϵ_4 allele. Hence, in an effort to elucidate potential links between ApoE genotypes and serum metabolites in AD, one could state the following research questions:

RQ: What are the mechanistic links between ApoE genotype and serum metabolites in AD?

- (1) What is the effect of the ApoE genotype on the expression of serum metabolites?
- (2) How discriminatory are serum metabolites among ApoE genotypes?
- (3) How do the network topologies of metabolites differ among ApoE genotypes?

1.3. Data. The data were collected from $n = 127$ AD patients with or without the ApoE4 allele, in the context of the Amsterdam Dementia Cohort [21], [32] and contain $p = 230$ metabolites (amines, organic acids, lipids and oxidative stress compounds). The methodology for the metabolomic analysis and ApoE genotyping can be found at Leeuw *et al.*'s Blood-based metabolic signatures in Alzheimer's Disease [21]: SMT1 . The data was cleaned as described in the same article. The resulting data set is high-dimensional, in the sense that it contains more variables than observations ($p > n$). Another particularity of the data is the covariance and collinearity of the variables. 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.

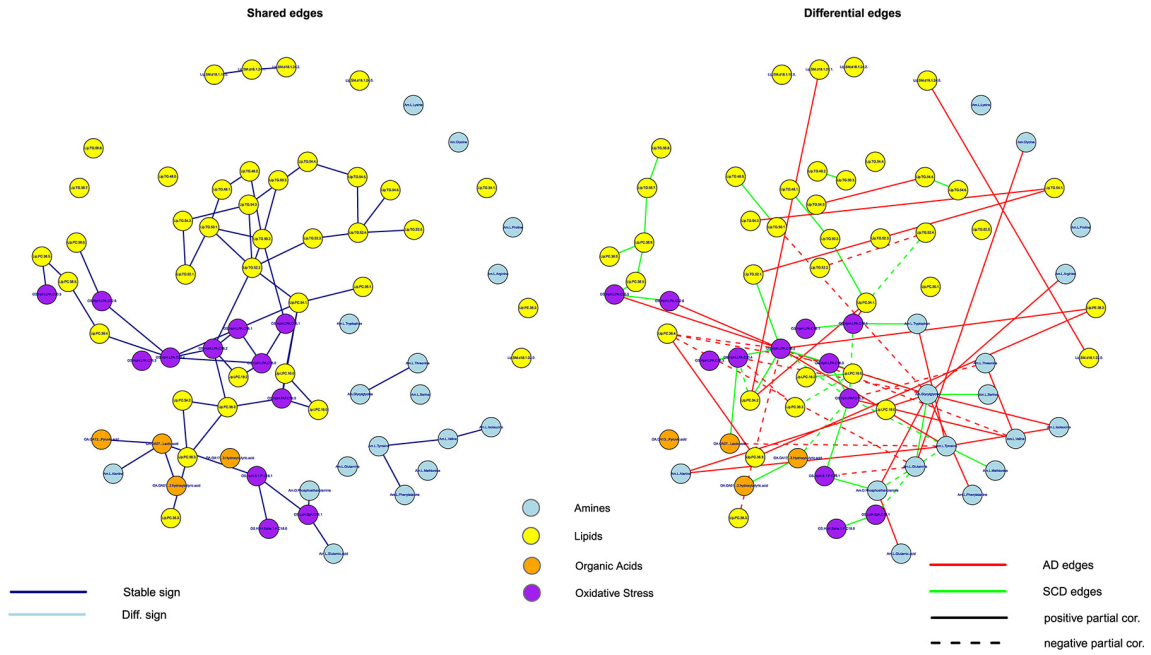


FIGURE 2. Mutual (left-hand panel) and distinct(right-hand panel) metabolic network topologies between ApoE ϵ_4 carriers and non-carriers in AD, as published by Leeuw *et al.* Red edges represent links that are present exclusively in ApoeE ϵ_4 carriers with AD. Green edges represent connections found in ApoE ϵ_4 non-carriers. Solid edges represent positive partial correlations, while dashed edges represent negative partial correlations. Abbreviations: SCD, subjective cognitive decline

2. METHODOLOGY

2.1. Data management. The FAIR principles for data management and stewardship in science were published by Wilkinson *et al.* in 2016 [33]. 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 will be considered at every step of the proposed project and implemented when applicable. All analysis will be performed in R, and the final Thesis report will be written in L^AT_EX, as is the current proposal.

2.2. Analysis. The degree to which a user can understand and interpret the prediction or decisions made by a statistical model is defined as *interpretability* [34]. It is of interest in this project to find the optimal balance between the performance of a model with 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 [35].

2.2.1. Differential expression. A straight-forward way to test whether the metabolites are differentially expressed between the 6 ApoE genotypes is ANOVA (Analysis of Variance) on p linear models. The dependent or response variable in each model would be a metabolite and the independent or explanatory variable would be the ApoE genotype. Let m_j represent the j -th metabolite and G_k the k -th possible ApoE genotype, an ANOVA model could be:

$$m_j = \mu + G_k + \varepsilon$$

Then, for every $j = 1, \dots, p$ the hypothesis test would be $H_0: G_k = 0$ for all $k \in \mathbb{N}, [1, 6]$ vs $H_a: \exists$ at least one $G_k \neq 0$. This implies p hypothesis tests, which creates the problem of multiple testing. A method to treat the latter is controlling False Discovery Rate [36], that is controlling the expected ratio of incorrectly rejected hypotheses, globally e.g. at an α of .05.

2.2.2. Classification. Considering interpretability, Multinomial Logistic Regression (MNL) is inherently interpretable. Let $y = k$, with $k \in N[1, 6]$ representing the k -th ApoE genotype (class) and β_{kj} its set of coefficients, β_{lj} the coefficients of the rest of classes for j -th metabolite, then an MNL model would calculate the probability

$$\Pr(y = k|X = x) = \frac{e^{\sum_{j=1}^p \beta_{kj}x_j}}{\sum_{l=1}^5 \sum_{j=1}^p e^{\beta_{lj}x_j}}$$

When $p > n$, the coefficient estimation method has low bias and high variance, in that small changes in the training data can result in very different coefficient estimates [37]. Regularization trades off a small increase in bias for a great decrease in variance, by shrinking the unimportant coefficients towards zero. LASSO (Least Absolute Shrinkage and Selection Operator) [38], also called L1-regularization shrinks the MNL coefficients to 0, thus weeding out spurious correlations and reducing the number of predictors [38]. It does so by introducing the term

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

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

A method to treat collinearity and high dimensionality is a 2-stage Maximum Likelihood(ML) factor analysis (FA), such as the one the package **FMradio** [30] performs. In the 1st stage, a L1-regularised ML estimation is used to filter out redundant features from the data matrix. In the second stage, ML FA projects the aforementioned matrix to an orthogonal space where the features are replaced by -fewer- factors that explain their covariance. One can then use the produced factor scores as predictors in MNL.

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 into branches and leaves(nodes) [39]. Random Forests (RF) are ensemble learning methods, that bag several DTs and average their decisions with

a majority vote. Despite RFs tend to outperform DTs, they often operate as a *black box* and are poorly interpretable.

The classification performance of the aforementioned models will be holistically assessed using repeated 10-fold CV-obtained AUC (Area Under the ROC curve) and other metrics such as Accuracy, Precision, Recall and F1-score.

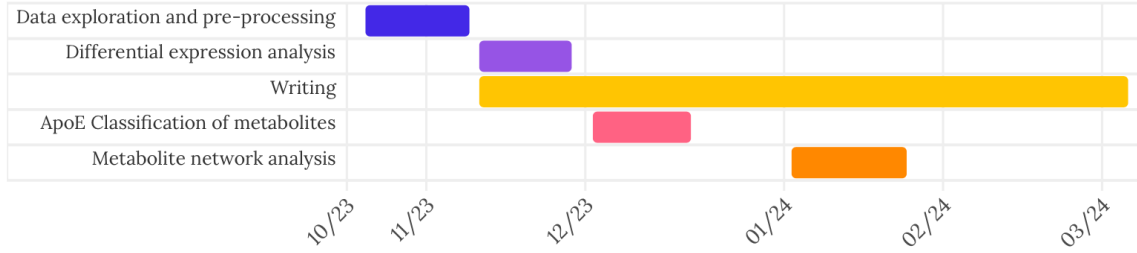
2.2.3. Network analysis. Network science offers a unifying framework for data and system representation, applicable to any domain [40]. A network, in an abstract sense, consists of nodes connected with links, also referred to as edges. In data science, a network whose nodes represent random features, whose joint probability distribution is defined by the ensemble of their edges is called *graphical model* [31]. A *Gaussian graphical model* (GGM) is an undirected graph that represents the conditional independence properties of the features [41]. For instance, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a GGM consisting of a set \mathcal{V} of p vertices, corresponding to random features Y_1, \dots, Y_p with joint probability distribution $P \sim N_p(\mathbf{0}, \Sigma)$, and set of edges \mathcal{E} , such that 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 Ω) mutually implies that the respective random features are independent, given the rest of features, which mutually implies that the corresponding features are not connected by an undirected edge ($(i, j) \notin \mathcal{E}$) [31].

Network analysis presents a unique approach to visualise high-dimensional and auto-correlated data. In this study, the package **rags2ridges** [31] may be used to generate the feature covariance matrix, as well as the precision matrix, regularise it and represent it in a GGM –as shown in Fig. ??.

3. PLANNING



REFERENCES

- [1] B. Penke, M. Szűcs, and F. Bogár, "New Pathways Identify Novel Drug Targets for the Prevention and Treatment of Alzheimer's Disease," *International Journal of Molecular Sciences*, vol. 24, no. 6, p. 5383, Mar. 2023, issn: 1422-0067. doi: 10.3390/ijms24065383. [Online]. Available: <https://www.mdpi.com/1422-0067/24/6/5383>.
- [2] A. Martin Prince, A. Wimo, M. Guerchet, *et al.*, *World Alzheimer Report 2015: The global impact of dementia: An analysis of prevalence, incidence, cost and trends*, Sep. 2015. [Online]. Available: <https://www.alzint.org/resource/world-alzheimer-report-2015/>.
- [3] A. Wimo, K. Seher, R. Cataldi, *et al.*, "The worldwide costs of dementia in 2019," *Alzheimer's & Dementia*, 2023, issn: 1552-5279. doi: 10.1002/ALZ.12901. [Online]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12901> <https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.12901> <https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12901>.
- [4] M. T. Heneka, M. J. Carson, J. E. Khoury, *et al.*, "Neuroinflammation in Alzheimer's disease," *The Lancet. Neurology*, vol. 14, no. 4, pp. 388–405, Apr. 2015, issn: 1474-4465. doi: 10.1016/S1474-4422(15)70016-5. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/25792098/>.
- [5] F. A. Edwards, "A Unifying Hypothesis for Alzheimer's Disease: From Plaques to Neurodegeneration," *Trends in Neurosciences*, vol. 42, no. 5, pp. 310–322, May 2019, issn: 1878108X. doi: 10.1016/j.tins.2019.03.003. [Online]. Available: <http://www.cell.com/article/S016622361930027X/fulltext> <http://www.cell.com/article/S016622361930027X/abstract> [https://www.cell.com/trends/neurosciences/abstract/S0166-2236\(19\)30027-X](https://www.cell.com/trends/neurosciences/abstract/S0166-2236(19)30027-X).
- [6] J. Rasmussen and H. Langerman, "Alzheimer's Disease – Why We Need Early Diagnosis," *Degenerative Neurological and Neuromuscular Disease*, vol. 9, p. 123, Dec. 2019. doi: 10.2147/DNND.S228939. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC6935598/>.
- [7] P. Scheltens, K. Blennow, M. M. Breteler, *et al.*, "Alzheimer's disease," *Lancet (London, England)*, vol. 388, no. 10043, pp. 505–517, Jul. 2016, issn: 1474-547X. doi: 10.1016/S0140-6736(15)01124-1. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26921134/>.
- [8] M. A. Beydoun, H. A. Beydoun, A. A. Gamaldo, A. Teel, A. B. Zonderman, and Y. Wang, "Epidemiologic studies of modifiable factors associated with cognition and dementia: Systematic review and meta-analysis," *BMC Public Health*, vol. 14, no. 1, pp. 1–33, Jun. 2014, issn: 14712458. doi: 10.1186/1471-2458-14-643/FIGURES/3. [Online]. Available: <https://bmcpublishing.biomedcentral.com/articles/10.1186/1471-2458-14-643>.
- [9] C. Van Cauwenberghe, C. Van Broeckhoven, and K. Sleegers, "The genetic landscape of Alzheimer disease: clinical implications and perspectives," *Genetics in Medicine* 18:5, vol. 18, no. 5, pp. 421–430, Aug. 2015, issn: 1530-0366. doi: 10.1038/gim.2015.117. [Online]. Available: <https://www.nature.com/articles/gim2015117>.
- [10] J. Hardy, "Alzheimer's disease: The amyloid cascade hypothesis: An update and reappraisal," *Journal of Alzheimer's Disease*, vol. 9, no. s3, pp. 151–153, Jan. 2006, issn: 1387-2877. doi: 10.3233/JAD-2006-9S317.
- [11] K. P. Kepp, N. K. Robakis, P. F. Høiland-Carlson, S. L. Sensi, and B. Vissel, "The amyloid cascade hypothesis: an updated critical review," 2023. doi: 10.1093/brain/awad159. [Online]. Available: <https://doi.org/10.1093/brain/awad159>.
- [12] M. Kurkinen, M. Fuek, K. Fuek, J. A. Beszej, D. Kurpas, and J. Leszek, "The Amyloid Cascade Hypothesis in Alzheimer's Disease: Should We Change Our Thinking?" *Biomolecules* 2023, Vol. 13, Page 453, vol. 13, no. 3, p. 453, Mar. 2023, issn: 2218-273X. doi: 10.3390/Biom13030453. [Online]. Available: <https://www.mdpi.com/2218-273X/13/3/453/html> <https://www.mdpi.com/2218-273X/13/3/453>.
- [13] D. K. Barupal, R. Baillie, S. Fan, *et al.*, "Sets of coregulated serum lipids are associated with Alzheimer's disease pathophysiology," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 11, pp. 619–627, Dec. 2019, issn: 2352-8729. doi: 10.1016/J.DADM.2019.07.002.
- [14] R. Fernández-Calle, S. C. Konings, J. Frontiñán-Rubio, *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, vol. 17, no. 1, pp. 1–47, Sep. 2022, issn: 1750-1326. doi: 10.1186/s13024-022-00566-4. [Online]. Available: <https://link.springer.com/articles/10.1186/s13024-022-00566-4> <https://link.springer.com/article/10.1186/s13024-022-00566-4>.
- [15] P. Proitsi, M. Kim, L. Whaley, *et al.*, "Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis," *Alzheimer's & Dementia*, vol. 13, no. 2, pp. 140–151, Feb. 2017, issn: 1552-5260. doi: 10.1016/J.JALZ.2016.08.003.
- [16] P. Oeckl and M. Otto, "A Review on MS-Based Blood Biomarkers for Alzheimer's Disease," *Neurology and Therapy*, vol. 8, no. 2, pp. 113–127, Dec. 2019, issn: 21936536. doi: 10.1007/s40120-019-00165-4/TABLES/3. [Online]. Available: <https://link.springer.com/article/10.1007/s40120-019-00165-4>.
- [17] M. A. Husain, B. Laurent, and M. Plourde, "APOE and Alzheimer's Disease: From Lipid Transport to Physiopathology and Therapeutics," *Frontiers in Neuroscience*, vol. 15, p. 630502, Feb. 2021, issn: 1662453X. doi: 10.3389/FNINS.2021.630502/BIBTEX.

- [18] C. Gan, Y. Zhang, F. Liang, X. Guo, and Z. Zhong, "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*, vol. 20, no. 1, pp. 1–12, Dec. 2022, issn: 14777819. doi: 10.1186/S12957-022-02748-2/FIGURES/2. [Online]. Available: <https://wjso.biomedcentral.com/articles/10.1186/s12957-022-02748-2>.
- [19] S. Parhizkar and D. M. Holtzman, "APOE mediated neuroinflammation and neurodegeneration in Alzheimer's disease," *Seminars in Immunology*, vol. 59, p. 101594, Jan. 2022, issn: 1044-5323. doi: 10.1016/J.SMIM.2022.101594.
- [20] M. Arnold, K. Nho, A. Kueider-Paisley, *et al.*, "Sex and APOE $\epsilon 4$ genotype modify the Alzheimer's disease serum metabolome," *Nature Communications* 2020 11:1, vol. 11, no. 1, pp. 1–12, Mar. 2020, issn: 2041-1723. doi: 10.1038/s41467-020-14959-w. [Online]. Available: <https://www.nature.com/articles/s41467-020-14959-w>.
- [21] F. A. de Leeuw, C. F. Peeters, M. I. Kester, *et al.*, "Blood-based metabolic signatures in Alzheimer's disease," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, vol. 8, pp. 196–207, Jan. 2017, issn: 2352-8729. doi: 10.1016/J.DADM.2017.07.006.
- [22] R. E. Green, J. Lord, M. A. Scelsi, *et al.*, "Investigating associations between blood metabolites, later life brain imaging measures, and genetic risk for Alzheimer's disease," *Alzheimer's Research and Therapy*, vol. 15, no. 1, pp. 1–13, Dec. 2023, issn: 17589193. doi: 10.1186/S13195-023-01184-Y/FIGURES/3. [Online]. Available: <https://alzres.biomedcentral.com/articles/10.1186/s13195-023-01184-y%20http://creativecommons.org/publicdomain/zero/1.0/>.
- [23] V. R. Varma, A. M. Oommen, S. Varma, *et al.*, "Brain and blood metabolite signatures of pathology and progression in Alzheimer disease: A targeted metabolomics study," *PLOS Medicine*, vol. 15, no. 1, e1002482, Jan. 2018, issn: 1549-1676. doi: 10.1371/JOURNAL.PMED.1002482. [Online]. Available: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002482>.
- [24] L. Sun, D. Guo, Y. Jia, *et al.*, "Association between Human Blood Metabolome and the Risk of Alzheimer's Disease," *Annals of Neurology*, vol. 92, no. 5, pp. 756–767, Nov. 2022, issn: 1531-8249. doi: 10.1002/ANA.26464. [Online]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1002/ana.26464%20https://onlinelibrary.wiley.com/doi/abs/10.1002/ana.26464%20https://onlinelibrary.wiley.com/doi/10.1002/ana.26464>.
- [25] L. Jia, J. Yang, M. Zhu, *et al.*, "A metabolite panel that differentiates Alzheimer's disease from other dementia types," *Alzheimer's & Dementia*, vol. 18, no. 7, pp. 1345–1356, Jul. 2022, issn: 1552-5279. doi: 10.1002/ALZ.12484. [Online]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1002/alz.12484%20https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.12484%20https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12484>.
- [26] Z. Huo, L. Yu, J. Yang, Y. Zhu, D. A. Bennett, and J. Zhao, "Brain and blood metabolome for Alzheimer's dementia: findings from a targeted metabolomics analysis," *Neurobiology of Aging*, vol. 86, pp. 123–133, Feb. 2020, issn: 0197-4580. doi: 10.1016/J.NEUROBIOLAGING.2019.10.014.
- [27] W. C. Weng, W. Y. Huang, H. Y. Tang, M. L. Cheng, and K. H. Chen, "The Differences of Serum Metabolites Between Patients With Early-Stage Alzheimer's Disease and Mild Cognitive Impairment," *Frontiers in Neurology*, vol. 10, p. 482026, Nov. 2019, issn: 16642295. doi: 10.3389/FNEUR.2019.01223/BIBTEX.
- [28] B. N. Simpson, M. Kim, Y. F. Chuang, *et al.*, "Blood metabolite markers of cognitive performance and brain function in aging," *Journal of Cerebral Blood Flow and Metabolism*, vol. 36, no. 7, pp. 1212–1223, Jul. 2016, issn: 15597016. doi: 10.1177/0271678X15611678/ASSET/IMAGES/LARGE/10.1177{_}0271678X15611678-FIG4.JPEG. [Online]. Available: <https://journals.sagepub.com/doi/full/10.1177/0271678X15611678>.
- [29] T. Oka, Y. Matsuzawa, M. Tsuneyoshi, and Y. Nakamura, "Multiomics analysis to explore blood metabolite biomarkers in an Alzheimer's Disease Neuroimaging Initiative cohort," Jun. 2023. doi: 10.21203/RS.3.RS-2973576/V1. [Online]. Available: <https://www.researchsquare.com%20https://www.researchsquare.com/article/rs-2973576/v1>.
- [30] C. F. W. Peeters, C. Übelhör, S. W. Mes, *et al.*, "Stable prediction with radiomics data," Mar. 2019. [Online]. Available: <https://arxiv.org/abs/1903.11696v1>.
- [31] C. F. Peeters, A. E. Bilgrau, and W. N. van Wieringen, "rags2ridges: A One-Stop-2-Shop for Graphical Modeling of High-Dimensional Precision Matrices," *Journal of Statistical Software*, vol. 102, pp. 1–32, May 2022, issn: 1548-7660. doi: 10.18637/JSS.V102.I04. [Online]. Available: <https://www.jstatsoft.org/index.php/jss/article/view/v102i04>.
- [32] W. M. Van Der Flier and P. Scheltens, "Amsterdam Dementia Cohort: Performing Research to Optimize Care," *Journal of Alzheimer's disease : JAD*, vol. 62, no. 3, pp. 1091–1111, 2018, issn: 1875-8908. doi: 10.3233/JAD-170850. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/29562540/>.
- [33] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data* 2016 3:1, vol. 3, no. 1, pp. 1–9, Mar. 2016, issn: 2052-4463. doi: 10.1038/sdata.2016.18. [Online]. Available: <https://www.nature.com/articles/sdata201618>.
- [34] R. Elshawi, M. H. Al-Mallah, and S. Sakr, "On the interpretability of machine learning-based model for predicting hypertension," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–32, Jul. 2019, issn: 14726947. doi: 10.1186/S12911-019-0874-0/FIGURES/48. [Online]. Available: <https://bmcmidinformedecismak.biomedcentral.com/articles/10.1186/s12911-019-0874-0>.
- [35] S. Geman, E. Bienenstock, and R. Doursat, "Neural Networks and the Bias/Variance Dilemma," *Neural Computation*, vol. 4, no. 1, pp. 1–58, Jan. 1992, issn: 0899-7667. doi: 10.1162/NECO.1992.4.1.1. [Online]. Available: <https://dx.doi.org/10.1162/neco.1992.4.1.1>.

- [36] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995, issn: 00359246. doi: 10.2307/2346101.
- [37] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning with Applications in R Second Edition," 2023.
- [38] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, Jan. 1996, issn: 2517-6161. doi: 10.1111/j.2517-6161.1996.tb02080.x. [Online]. Available: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.2517-6161.1996.tb02080.x>
<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x>
<https://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1996.tb02080.x>.
- [39] Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Archives of Psychiatry*, vol. 27, no. 2, p. 130, Apr. 2015, issn: 10020829. doi: 10.11919/J.ISSN.1002-0829.215044. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/>
<https://pubmed.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/?report=abstract>
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4466856/>.
- [40] A.-L. Barabási, *Network Science*. Cambridge University Press, 2015. [Online]. Available: <http://networksciencebook.com/>.
- [41] D. Koller and N. Friedman, "Probabilistic Graphical Models: Principles and Techniques,"