**Proposal details**

Title of the project:
Using interpretable machine learning to diagnose Alzheimer's Disease through blood-based immunological biomarkers

Duration of the project (in months):
6

**Public summary**

English public summary (96/100):
*This research seeks to improve Alzheimer's Disease (AD) diagnosis through identifying blood-based immunological biomarkers using interpretable machine learning. With rising AD prevalence straining Dutch healthcare, current diagnostic methods are either invasive or expensive. Utilising the body's neuroinflammatory responses, we will use LC-MS to identify and ELISA to validate potential biomarkers. A Logistic Regression machine learning model will be developed for AD diagnosis using these potential biomarkers, benchmarked against established ones. The goal is a cost-effective, minimally invasive diagnostic tool, which would enhance our understanding of AD, thus alleviating the societal and economic burdens of the disease.*

**Budget**

| Type of costs | Clarification | Costs in Euro |
|---|---|---|
| Personnel | Nurse[1] | 10,000 |
| | Doctor[2] | 10,000 |
| | Bioinformatician to develop machine learning model[3] | 0 |
| | PI or PhD or master student[3] | 0 |
| | Lab technician specialised in LC-mass spectrometry and ELISA[3] | 0 |
| Material | Syringes (€20 per individual) | 10,000 |
| | Needles (€20 per 100 pieces) | 100 |
| | LC-Mass spectrometry (€180 per three hours = approx. 36 samples)[4] | 2,500 |
| | ELISA (€746 per assay = approx. 40 samples in duplicate)[5] | 9,325 |
| | Disposables | 7,975 |
| Computing | Cloud computing rental for model training | 100 |
| **Total request from NWO** | | **50,000** |

**Budget justification (86/150)**

1. We intend to employ a nurse for six months for the collection of blood samples.
2. We intend to employ a doctor for two months to interview and diagnose patients.
3. They will be compensated by a lab we will be working with in collaboration.
4. LC-MS is not specific, allowing us to pick the best molecules to investigate. We need 500 samples to train our model accordingly.
5. ELISA can be used for verification when we have chosen our molecules of interest through LC-MS.

**Project proposal**

**Alzheimer's Disease - a public health issue:** We are interested in Alzheimer's Disease (AD) because it not only affects the diagnosed individuals themselves but also their close ones, clinicians, and all of society[1]. As the prevalence and healthcare costs of AD are predicted to double by 2050, it is expected that about 300,000 Dutch citizens will have AD and the Dutch healthcare system will be facing costs of 15.6 billion euro[2,3].
AD is a specific as well as the most common type of dementia[3]. Dementia is an umbrella term that describes a group of diseases that share a particular set of symptoms, entailing problems in memory, language, problem-solving, and other thinking processes[3] – all being building blocks of navigating through everyday life. AD is characterized by I) abnormal accumulation of the proteins amyloid-β42, resulting in the formation of amyloid plaques, and tau-protein, leading to tau-tangles inside of neurons[4] II) neuronal degeneration across several parts of the brain, which enable cognitive functions like memory[3]. It has been shown, that neuroinflammation plays a central role in the development of AD[5,6]. The state of neuroinflammation can be assessed by measuring molecules, or biomarkers, in the blood. Biomarkers are biological changes that can be measured to, for instance, indicate the absence or presence of a disease.
**The critical need for advancements of Alzheimer's diagnosis:** At the moment, clinicians measure either the state of tau or amyloid-β42 accumulation or neural degeneration (atrophy) to diagnose AD[3]. A sample of cerebrospinal fluid (CSF) can be indicative of the amyloid-β42 and tau protein levels but requires a painful lumbar puncture, which includes inserting a special needle between two vertebrae in the back[3,7]. This method carries risks and the most common side effects include local pain and severe headaches. Alternatively, using a brain imaging technique called Positron Emission Tomography (PET) can visualise the amount of tau and amyloid-β42 accumulation[3]. However, PET is quite expensive and entails the injection of a radioactive agent to make the proteins visible[7]. As previously mentioned, using blood-based biomarkers would not entail similar expenses or discomfort. It has been established that blood can be useful in detecting AD, but more accurate and sensitive blood-based biomarkers are needed for reliable diagnosis[6].
**Transforming AD diagnosis - our idea of an AI-powered tool:** The *Dutch national dementia strategy 2021-2030* views artificial intelligence (AI) as a key tool to accelerate the state of the art and calls for innovative, scientific solutions to prevent and combat expected dementia diagnoses[8]. **With our research we aim to improve the current state of diagnosing AD by offering an affordable, interpretable, and minimally invasive diagnostic machine learning tool that is based on blood samples.** This approach not only aligns with the Dutch national dementia strategy but also promises to reduce the costs associated with AD, as well as to make the diagnostic process less invasive, as taking blood samples is both less expensive and invasive than contemporary diagnostic tools.

**Research Strategy**

**Objective 1:** *Identify immunological components that can be used to detect AD using blood samples.* The blood samples will be taken from healthy patients as a control, and patients who have AD. We are going to use 250 samples per cohort, 500 total. We plan to use liquid chromatography-mass spectrometry (LC-MS) to identify immunological molecules that have an association with AD. Mass-spectrometry (MS) is an analytical technique that is used to detect biomolecules (like peptides) and small molecules (like metabolites) that are present in the blood samples. It determines the presence of particles based on their mass-to-charge ratio[9]. The addition of liquid chromatography (LC) allows liquid samples high in protein concentrations (e.g., blood plasma) to be analysed, as normal MS is incapable of doing so, allowing us to detect immunological components[10,11]. We will filter the results to only include the immunological molecules and we will subsequently calculate their concentration. This method differs from CSF analysis, which uses targeted Enzyme-Linked Immuno Sorbent Assay (ELISA) measurements to measure the concentration of established biomarkers.
**Objective 2:** *Perform ELISA validation on the samples.* We want to ensure that LC-MS does not give false positive results and, thus, result in an incorrectly trained model. We will, therefore, use ELISA for validation of the LC-MS results, as ELISA is a much more specific technique than LC-MS.
**Objective 3:** *Create a supervised machine learning model for AD diagnosis, using the data validated in objective 2.* We will make use of a Logistic Regression (LR) model due to its several advantageous attributes. LR is interpretable, which allows clinicians and researchers to comprehend the model's decision-making process, making it transparent and trustworthy. It gives probability-based results, which can be crucial for borderline cases. The coefficients derived from LR allow for an understanding of the importance and influence of each immunological component, which is valuable for identifying the most

significant biomarkers in AD. Finally, LR is computationally efficient, an essential aspect in medical environments where both time and expense are significant considerations.
To evaluate the performance of model based on our found immunological components, it will be benchmarked against a LR model that relies on well-known biomarkers, namely tau and amyloid-β42[4]. We will employ several relevant metrics to ensure a well-rounded evaluation of the model's diagnostic performance. The concordance index (C-index) will serve as a key indicator of the model's discriminative ability, essential for prioritizing high-risk patients, while the Root Mean Square Error (RMSE) will offer a straightforward evaluation of the model's accuracy in predicting disease probability. We will also compute sensitivity and specificity to capture the model's reliability in correctly identifying affected individuals and avoiding false alarms, which is critical to patient care and resource allocation.
In the final step, we will assess the importance of individual immunological markers in AD diagnosis by examining the coefficients of our LR model. This analysis will identify which immunological markers are the most significant in the context of AD diagnosis, guiding future research and potential clinical applications.

**Risks and Challenges:** The effectiveness of our machine learning model for diagnosing AD is dependent on the quality and quantity of the training data we collect. Any biases in the data could be reinforced by the model, which could lead to more unequal healthcare outcomes, such as misdiagnoses or systemic inaccuracies that disproportionately affect certain demographic groups. To mitigate this, our strategy emphasizes gathering a substantial dataset that is accurate, representative, and diverse, by recruiting a heterogeneous sample.
A potential challenge in our approach is the ability of LR to accurately capture the intricate complexities associated with AD diagnosis. While this model offers clarity and interpretability, its simplicity might limit performance. If necessary, we might consider transitioning to more advanced models like deep learning to enhance accuracy.
**Knowledge Utilization:** The link between the immune system and AD is an active area of research, that has gained attention[12]. With our project, we add to this new field by investigating the potential of using blood-based immunological biomarkers. By creating an interpretable, less invasive, and cost-effective diagnostic machine learning model, we hope to offer a useful tool for clinicians to diagnose AD and lessen costs that burden our public health care system. This model can also show healthcare professionals how each biomarker affects the diagnosis to increase trust. In addition, AI systems can learn and adapt over time. As more data becomes available, these systems can continuously improve their diagnostic accuracy, staying up to date with the latest research and findings. Beyond improving the diagnostic for AD, this project may be a pioneering proof of concept for the utilization of machine learning in biomarker discovery. Also, our project offers a novel perspective on AD by providing precious ex vivo data, as we use human blood samples. Lastly, while learning more about the connection between these biomarkers and AD, new treatments might be inspired, as discovering specific immunological biomarkers may offer crucial insights into the underlying mechanisms of the disease.

**References**


1. Nandi, A., Counts, N. Z., Chen, S., Seligman, B., Tortorice, D. L., Vigo, D. E., & Bloom, D. E. (2022). Global and regional projections of the economic burden of Alzheimer's disease and related dementias from 2019 to 2050: A value of statistical life approach. EClinicalMedicine, 51, 101580. https://doi.org/10.1016/j.eclinm.2022.101580
2. Prevalence of dementia in Europe. (n.d.). Alzheimer Europe. https://www.alzheimer-europe.org/dementia/prevalence-dementia-europe
3. 2023 Alzheimer's disease facts and figures. (2023). Alzheimer's & Dementia, 19(4), 1598–1695. https://doi.org/10.1002/alz.13016
4. Teunissen, C. E., Chiu, M., Yang, C., Yang, S., Scheltens, P., Zetterberg, H., & Blennow, K. (2018). Plasma amyloid-B (AB42) correlates with cerebrospinal fluid AB42 in Alzheimer's disease. Journal of Alzheimer's Disease, 62(4), 1857–1863. https://doi.org/10.3233/jad-170784
5. Heneka, M. T., Carson, M. J., Khoury, J. E., Landreth, G. E., Brosseron, F., Feinstein, D. L., Jacobs, A. H., Wyss-Coray, T., Vitórica, J., Ransohoff, R. M., Herrup, K., Frautschy, S. A., Finsen, B., Brown, G. C., Verkhratsky, A., Yamanaka, K., Koistinaho, J., Latz, E., Halle, A., . . . Kummer, M. P. (2015). Neuroinflammation in Alzheimer's disease. Lancet Neurology, 14(4), 388–405. https://doi.org/10.1016/s1474-4422(15)70016-5
6. Angiulli, F., Conti, E., Zoia, C. P., Re, F., Appollonio, I., Ferrarese, C., & Tremolizzo, L. (2021). Blood-Based biomarkers of neuroinflammation in Alzheimer's Disease: a central role for periphery? Diagnostics, 11(9), 1525. https://doi.org/10.3390/diagnostics11091525
7. Mason, E. J., Donahue, M. J., & Ally, B. A. (2013). Using magnetic resonance imaging in the early detection of Alzheimer's disease. In InTech eBooks. https://doi.org/10.5772/54445
8. Ministerie van Algemene Zaken. (2020). National Dementia Strategy 2021-2030. Publication. Government.nl. https://www.government.nl/documents/publications/2020/11/30/national-dementia-strategy-2021-2030
9. Murayama, C., Kimura, Y., & Setou, M. (2009). Imaging mass spectrometry: principle and application. Biophysical Reviews, 1(3), 131–139. https://doi.org/10.1007/s12551-009-0015-6
10. Pitt, J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. PubMed, 30(1), 19–34. https://pubmed.ncbi.nlm.nih.gov/19224008
11. Robotham, A., & Kelly, J. F. (2020). LC-MS characterization of antibody-based therapeutics. In Elsevier eBooks (pp. 1–33). https://doi.org/10.1016/b978-0-08-103019-6.00001-1
12. Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., & Jones, E. (2011). Alzheimer's disease. The Lancet, 377(9770), 1019–1031. https://doi.org/10.1016/s0140-6736(10)61349-9