AI meets MD: Combining Machine learning algorithms and medical experts' diagnosis (MD) for new beginnings in the prediction of Alzheimer's disease progression.

Oliver Bruton, Sumbul Jafri, Raxide Andrade Leon

Supervisor: Prof. Dr. Andrea Hildebrandt

Alzheimer's Disease (AD) is becoming a significant economic and social burden, yet the conventional process of medical diagnosis (MD) evinces significant proneness to error. One way of improving diagnostic accuracy may lie in data analysis performed by machine learning algorithms (MLA). However, limited availability of longitudinal biomarker information of AD progression leads to a data imbalance problem. This issue may be alleviated by designing Clinical Decision Support Systems (CDSS) incorporating both algorithmic and clinical predictions. Thus, the aim of the current study was to investigate the concordance between the assessment of progression predictors by clinicians and MLA. A survey including plots depicting sampling distributions of the most relevant predictors for three AD progression groups was created. Nine clinicians were contacted and asked to assign each patient to one of the groups. None of the clinicians fully completed the survey and only one provided detailed feedback. The results are interpreted as indicative of health professionals' scepticism towards MLA.

Clinical decision support systems, CDSS, machine learning, Alzheimer's disease

Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disorder, which is most prominently characterised by deficits in cognitive and specifically mnestic functions (De Ture & Dickson, 2019). Approximately 50-70% of all dementia cases are thought to be attributable to the disease (van Praag, 2018). Especially in the west, the adverse effects of dementia are thought to have a severe impact, contingent on shifts in demographics and over-ageing (Banerjee, 2012; Maia & De Mendoça, 2002; Niccoli & Partridge, 2012). This development is likely to be further exacerbated by a lack of qualified caretaking personnel described for countries such as Germany and Japan (Waldenberger, Naegele, Kudo, & Matsuda, 2022). Reports concerning the projected financial consequences of AD and related forms of dementia are alarming. While estimated at around US$818 billion in 2015, the global costs could rise to as much as US$2 trillion by 2030 (El-Hayek et al., 2019, Wimo et al., 2017). Consequentially, the need to effectively employ and adequately distribute relevant (monetary) means has been described as one of the greatest challenges for modern medicine and health politics (Mohelska, Maresova, Valis, & Kuca, 2015). The process of medical diagnosis (MD) is one of the crucial aspects of determining the allocation of resources to different patient groups (Meaney, Croke, & Kirby, 2005). Evidently, the conventional form of arriving at a diagnosis relies upon the experience and judgement of a qualified medical professional. However, this process has repeatedly been shown to be susceptible to human error, owing to e.g., cognitive biases (Abimanyi-Ochom et al., 2019; Edmonds et al., 2016, Saposnik, Redelmeier, Ruff, & Tobler, 2016). One way of dealing with the shortcomings in traditional clinical decision-making may lie in the incorporation of machine learning (ML) algorithms, the implementation of which has been previously discussed (Kohli & Arora, 2018; Richens, Lee, & Johri, 2020). This proposition is predicated on the well-known abundance of data in the health sector (Bennett & Hauser, 2013). In the context of AD, previous studies have indicated that MLA can indeed serve the purpose of assisting in the identification of the disease (Chang, Lin, & Lane, 2021; Trambaioili, Lorena, Fraga, Kanda, Anghinah, & Nitrini, 2011).

In reference to the domain of AD progression, Andrade Leon, Bruton, & Jafri, 2021, have demonstrated that the MLA available to date are subject to limited availability of longitudinal biomarker data pertaining to patients at high risk for fast progression of AD. This means that available open access AD data, which can be deployed for the creation of prediction models of AD development, are afflicted with the problem of class imbalance between various forms of progression (i.e., slow, moderate, and fast). This inherent deficiency may represent a significant hindrance to the design of practically valid CDSS. However, issues such as a dearth of appropriate data may be counteracted by combining algorithmic analysis with expertise-based judgements from experienced clinicians (Xie, Niu, Liu, Chen, Tang, & Yu, 2021). By applying different MLA to a large-scale study of AD (ADNI, Jack et al., 2008), Andrade Leon, Bruton, & Jafri (2021) were able to identify the most relevant biomarker predictors of AD progression, which included clinical, demographic and imaging data. The objective of the current study is to use this information and investigate the degree to which experienced health professionals versed in the area of dementia agree with the algorithmic predictions and their evaluation thereof. This data will be then used to improve the previous predictions, which is thought to facilitate the utility of CDSS for the prognostication of dementia progression severity.

Methods and Procedure

For assessing the concordance of AD progression prediction between MLA and health practitioners, a survey was devised which was based on the 10 most relevant predictors previously identified by analysing data from the ADNI-database (for further details in this respect see Andrade Leon, Bruton, & Jafri, 2021). The list of relevant features consisted of:

- chronological baseline age,
- score on the Rey Auditory Verbal Learning test (RAVLT; range 0-15)
- functional assessment questionnaire (range 0-30)
- baseline biomarker amyloid-ß peptide (variant Aß1-42; plasma concentration pg/ml)
- baseline biomarker batch analysis of CSF biomarkers TAU (pg/ml)

- baseline MRI whole brain volume (ml)
- baseline MRI ventricular volume (ml)
- longitudinal MRI measure 1.1: cumulative temporal lobe atrophy; change score from timepoint 0 to 1
- longitudinal MRI measure 1.2: cumulative temporal lobe atrophy; change score from timepoint 0 to 2
- longitudinal MRI measure 2: average within an anatomically defined region of interest including bilateral temporal lobes; change score from timepoint 0 to 1

The survey depicted the sampling distributions of those predictors for the three progression groups (slow, intermediate, and fast) along with their mean values. A total of 30 patients were located and marked within said distributions. The survey was conducted online using the Soscisurvey software (Leiner, 2019).

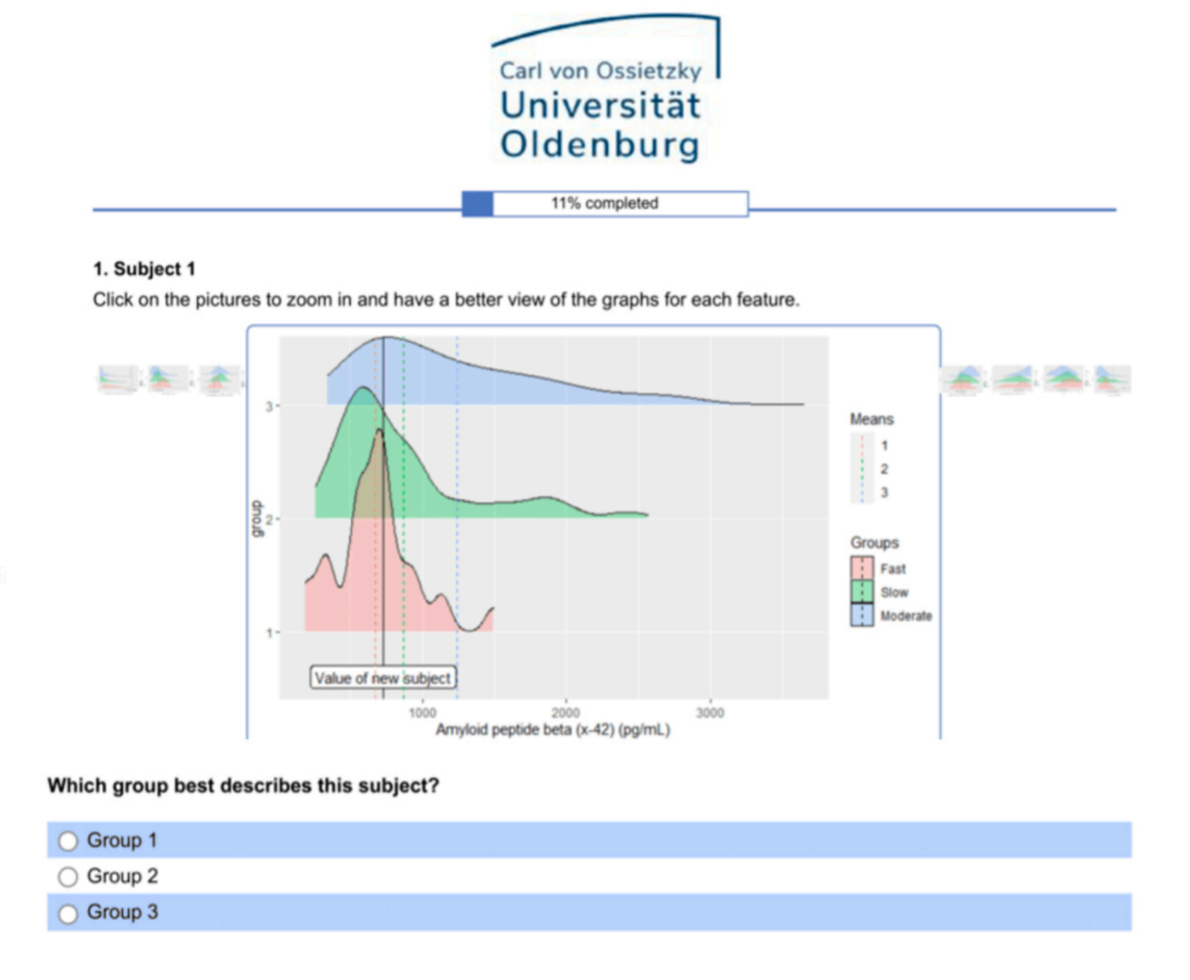

Fig.1: Example of a slide as used in the employed survey. The image shows a specific predictor's value for subject 1, along with this predictor's distributions and means of the three progression groups. The clinician's task is to assess which group the respective subject should best be allotted to.

The task for any participating clinician was to inspect each of the depicted plots and, based on the sampling distributions of the progression groups and the clinician's medical expertise, to decide, which of the progression groups the respective patient was most likely to belong to. After construction of the survey, we contacted 9 experienced clinicians from various domains to capture many different disciplinary perspectives (psychiatry, geriatrics, neuropsychology, and neurosurgery). The participants were recruited anonymously from Germany and Mexico.

Results

Although the practitioners partaking attempted and partially completed the survey, none of them fully completed it. However, one clinician provided detailed feedback. He stated that he was unable to answer the survey questions as he rarely followed up the patients he was working with – an essential prerequisite for judging disease progression. Moreover, he provided a critical reflection of the employed instrument as well as the implementation of MLA for diagnostic purposes in general. He pointed out, that it may be relevant to consider the aspect of disease onset. Additionally, he referred to non-specified evidence which suggested that some of the predictors previously identified as relevant by MLA were in fact practically irrelevant (e.g., atrophy of the medio-basal temporal lobe). Moreover, he stressed the perceived practical lack of utility regarding imaging-based diagnostics, as the required technology was not readily available for widespread usage. Eventually, he voiced his doubts concerning the applicability of the features for progression prediction in the context of individual cases. He noted that, according to his experience, degree of AD severity may be compensated by external factors such as adequate patient support at home.

Discussion

The current study aimed to assess the agreement between MLA-based predictions of AD progression on the one hand and the experience-informed judgement by practicing clinicians on the other. The results of this exploratory investigation are construed as indicating a significant sense of scepticism of medical professionals towards the utilisation of predictions derived from MLA. This scepticism has been previously described and discussed (e.g., Sources). A possible reason for this scepticism may lie in the fact that the scientific evidence allocated by CDSS is capable of defying firmly established conceptions in the medical practice (Laka, Milazzo, & Merlin, 2021; Liberati et al., 2017). Particularly, clinicians may have the notion that computerised CDSS may lead to a reduction of their professional autonomy, challenge longstanding hierarchical structures or serve as legal evidence against the clinician in the context of controversy (Liberati et al., 2017). Therefore, future studies aiming at the inclusion of algorithmic information for clinical decision-making should pay close attention to this aspect. As MLA are statistical tools, the appropriate description of the strengths and potential weaknesses/shortcomings of these instruments may decrease the resistance and doubtful perceptions of health practitioners towards CDSS. Despite these limitations however, ML-based analysis has proven beneficial for application in the realm of medicine, which further highlights the necessity of overcoming the degree of reserve that may exist in medicine.

References

Abimanyi-Ochom, J., Bohingamu Mudiyanselage, S., Catchpool, M., Firipis, M., Wanni Arachchige Dona, S. & Watts, J. J. (2019). Strategies to reduce diagnostic errors: a systematic review. BMC medical informatics and decision making, 19(1), pp. 1-14.

Andrade Leon, R., Bruton, O.J. & Jafri, S. (2021, November). AI meets AD: Coining the future of early and accurate prediction of Alzheimer’s Disease progression. Poster presentation as part of the TdLL at the University of Oldenburg, Germany.

Banerjee, S. (2012). The macroeconomics of dementia—will the world economy get Alzheimer's disease? Archives of medical research, 43(8), pp. 705-709.

Bennett, C. C. & Hauser, K. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. Artificial intelligence in medicine, 57(1), pp. 9-19.

Chang, C. H., Lin, C. H. & Lane, H. Y. (2021). Machine learning and novel biomarkers for the diagnosis of Alzheimer’s disease. International Journal of Molecular Sciences, 22(5), p. 2761.

DeTure, M. A. & Dickson, D. W. (2019). The neuropathological diagnosis of Alzheimer’s disease. Molecular neurodegeneration, 14(1), pp. 1-18.

Edmonds, E. C., Delano-Wood, L., Jak, A. J., Galasko, D. R., Salmon, D. P., Bondi, M. W. et al. (2016). “Missed” mild cognitive impairment: High false-negative error rate based on conventional diagnostic criteria. Journal of Alzheimer's Disease, 52(2), pp. 685-691.

El-Hayek, Y. H., Wiley, R. E., Khoury, C. P., Daya, R. P., Ballard, C., Evans, A. R. et al. (2019). Tip of the iceberg: assessing the global socioeconomic costs of Alzheimer’s disease and related dementias and strategic implications for stakeholders. Journal of Alzheimer's Disease, 70(2), pp. 323-341.

Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D. et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 27(4), pp. 685-691.

Kohli, P. S. & Arora, S. (2018). Application of machine learning in disease prediction. In 2018 4th International conference on computing communication and automation (ICCCA) (pp. 1-4). IEEE.

Laka, M., Milazzo, A. & Merlin, T. (2021). Factors that impact the adoption of clinical decision support systems (cdss) for antibiotic management. International journal of environmental research and public health, 18(4), 1901.

Leiner, D. J. (2019). SoSci Survey (Version 3.1.06) [Computer software]. Available at https://www.soscisurvey.de

Liberati, E. G., Ruggiero, F., Galuppo, L., Gorli, M., González-Lorenzo, M., Maraldi, M. et al. (2017). What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implementation Science, 12(1), pp. 1-13.

Maia, L. & De Mendonça, A. (2002). Does caffeine intake protect from Alzheimer's disease? European journal of neurology, 9(4), pp. 377-382.

Meaney, A. M., Croke, M. & Kirby, M. (2005). Needs assessment in dementia. International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences, 20(4), 322-329.

Mohelska, H., Maresova, P., Valis, M. & Kuca, K. (2015). Alzheimer's disease and its treatment costs: case study in the Czech Republic. Neuropsychiatric Disease and Treatment, 11, 2349.

Niccoli, T. & Partridge, L. (2012). Ageing as a risk factor for disease. Current biology, 22(17), pp. R741-R752.

Richens, J. G., Lee, C. M. & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. Nature communications, 11(1), 1-9.

Saposnik, G., Redelmeier, D., Ruff, C. C. & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. BMC medical informatics and decision making, 16(1), pp. 1-14.

Trambaiolli, L. R., Lorena, A. C., Fraga, F. J., Kanda, P. A., Anghinah, R. & Nitrini, R. (2011). Improving Alzheimer's disease diagnosis with machine learning techniques. Clinical EEG and neuroscience, 42(3), pp. 160-165.

van Praag, H. (2018). Lifestyle factors and Alzheimer's disease. Brain Plasticity, 4(1), 1.

Waldenberger, F., Naegele, G., Kudo, H. & Matsuda, T. (2022). Einleitung zu Alterung und Pflege als kommunale Aufgabe. In F. Waldenberger, G. Naegele, H. Kudo, T. Matsuda (Eds.) Alterung und Pflege als kommunale Aufgabe (pp. 85-100). Springer VS, Wiesbaden.

Wimo A., Guerchet M., Ali G.-C., Wu Y.-T., Prina A.M. & Winblad B. et al. (2017). The worldwide costs of dementia 2015 and comparisons with 2010. Alzheimers Dement 13, pp. 1–7.

Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S. & Yu, S. (2021). A survey on incorporating domain knowledge into deep learning for medical image analysis. Medical Image Analysis, 69, 101985