Assignment 1: Review of a journal article on the use of a machine learning or AI application for health or medical data 1-page review

Article

Ficiara, E., Boschi, S., Ansari, S., et al. *Machine Learning Profiling of Alzheimer's Disease Patients Based on Current Cerebrospinal Fluid Markers and Iron Content in Biofluids*. Front. Aging Neurosci., 21 February 2021 Sec. Alzheimer's Disease and Related Dementias. Volume 13 - 2021. https://doi.org/10.3389/fnagi.2021.607858

Health/medical topic or research area of the article

The mechanisms behind neurodegenerative diseases such as Alzheimer's Disease (AD) are not well understood due to the complexity of the human brain; hence many neurodegenerative diseases are considered chronic and incurable. AD was chosen for analysis due to its high prevalence, particularly among those over 65, and its significant negative impact on quality of life.

Research objective

This study uses machine learning (ML) to analyze iron levels in cerebrospinal fluid (CSF) along with biomarkers Abeta and tau to compare their correlations and diagnostic power for Alzheimer's disease (AD).

Analysis approach

The study focuses on measuring the effect of iron on the biomarkers used to diagnose AD (Abeta and tau peptides) as high levels of iron are linked to AD. Iron is an important molecule to the human body but becomes toxic if not regulated properly, thereby making it a possible early signaler of AD.

Patients were studied for iron-CSF levels with respect to CSF biomarkers and demographic traits. The dataset was processed using support vector machines (SVM) and logistic regression (LR) for multiclass classification in order to match patients to sample groups (ex. AD or control) based on patterns in their risk factors (ex Abeta-42 levels). This tests the diagnostic power of risk factors and identifies correlations between them.

Dataset description (4 marks)

Demographic (gender, age, and age-onset) and CSF (Abeta-42, p-tau, t-tau, MMSE, and s-Tf) data was collected from 69 hospital patients whom were divided into 4 groups: control (CT) (n=14), mild cognitive impairment (MCI) (n=17), AD (n=16), and fronto-temperal dementia (FTD) (n=22) via diagnostic and exclusion criteria. CSF samples were processed using spectrometry and results are made available in the supplementary material. The data was standardized (via Z-score), then grouped based on similarity and finally stratified into 2 datasets for training and testing of ML models.

Results/Findings

- Iron-CSF concentration was significantly higher in the AD group compared to all 3 other groups (CT, MCI, FTD). There is no significant difference between those 3 groups
- Abeta-42 had the highest diagnostic power as a biomarker in signaling AD, compared to p-tau

Discussion

This study was motivated by similar research that suggested a connection between iron and AD biomarkers. The incorporation of ML was also inspired by other research papers that used classification and clustering in the analysis of early-stage AD. Although the paper strengthened this research, it does not expand on the mechanisms behind iron and AD. The authors suggest a longitudinal study could shed more insight on how iron behaves at different stages of AD.

In terms of bias, the authors seem very insistent on proving their hypothesis that iron-CSF has strong diagnostic power for AD. Their selection of experimental groups is limited to only 4 cases despite high iron levels potentially indicating a multitude of other diseases; thus bringing external validity to question.

Conclusion

The research problem is unique and shows promise, but I question whether the data processing was over-engineered as the feature vectors were fairly small and could have been processed using statistical analysis rather than machine learning.