Peer Feedback

Report title: "Predicting Alzheimer And Cognitive Impairment From Handwriting Tasks"

For: M. van Dam, J. Koop and M. Urbanski

By: F.P. Meijvis, A.E. Maranus and W.B. Eiling

Summary Paragraph

In this partial reproduction of Cilia et al. (2022), the draft report employs a dataset with features related to handwriting to predict whether participants have Alzheimer's disease. Two different dimension reduction techniques are employed: (1) t-tests of the differences between patients and control for continuous predictors (threshold of $p < .001$) and (2) principal component analysis. After applying either method, a logistic regression model was fitted where the remaining variables or principal components predicted disease status. To determine which approach was superior, five-fold cross-validation will be performed, where the converged model with the smallest estimated error rate was selected.

Main Comments

1) General

Overall, we like the approach of predicting Alzheimer's patients based on the handwriting data. We have some small remarks titles of (sub)sections in the paper, since the introduction subtitles are quite long and could be improved in a way that they are more professionally stated and we would suggest subsections for the methods section to make it more structured.

2) Introduction

The introduction provides adequate context, but could make some links more explicit. For instance, it states that *"there is no standard way to collect data"*. Whereas this could be imagined to mean that there is no consensus of what the most valid and reliable writing tasks and features are, this is not made explicit. The dataset could also be described in a bit more detail. What do the data entries look like, are the variables numbers, categories, and what do we consider "high" or "low" valued data entries?

Furthermore, when looking at the approach by Cilia et al (2022), it is not explained what kind of classifiers they use. This makes the text a bit confusing, as it becomes hard to understand what the process exactly is.

3) Methods

In general, the paper could use more references when describing and motivating the two chosen methods. The authors don't really elaborate on why the benchmark method, the t-test, was chosen. We guess that it replicates the method of Cilia et al (2022), but could use referencing to this article.

Regarding the PCA, too much focus is put into explaining how it works. The draft report states that both the eigenvalue greater than 1 and elbow criteria are employed for the selection of the principal components. It is not stated when which method gains preference over the other. When applying the PCA to make predictions, it would seem to make more sense to employ the predictive accuracy as a criterion on which to base the number of components, to avoid reliance on the assumption "the directions in which the predictors show the most variation are the directions that are associated with Y (James et al., 2013)".

4) Results & Conclusion

Regarding the comparison of model performances, it is important to think about what parameters you find most important for choosing your best model. Given the context of your data, is it more important to correctly classify people as healthy or is it more important to correctly classify patients with the disease? This influences whether you look at specificity or sensitivity performance.

Suggested improvements:

- Explain technical features of the data in introduction section 1.1. Is the data numerical, categorical, what is a high-value or a low value?
- In the introduction section 1.2, explain what the classifiers of Cilia et. al (2022) are.
- Remove the description of how PCA works in the methods section, and put more focus on why PCA is applied and what decision criteria are implemented.
- Elaborate about what performance classifiers are important in choosing the best prediction model in this case for the diagnostics of Alzheimer's disease.