

Peer-Review

Xia, Matthew. 2026. "Predictive Overlap, Feature Reduction, and Robustness in Machine Learning-Based Cross-Sectional Classification of Alzheimer's Disease." *Journal of High School Science* 10 (2): 243–74. <https://doi.org/10.64336/001c.162155>.

I have several comments for this manuscript. In addition, point 1 represents a comment from the editor.

1. I am trying to see if you can get (any type) of regression when you plot redundancy versus (sensitivity, specificity, F1.... the usual metrics). (You do not right now). To do this, you will need a different way to compute redundancy. While we are reviewing your manuscript, can you compute the separate redundancy (for all the models) using the following methods:

2. VARIANCE of the top-10 features

3. MEAN PAIRWISE ABSOLUTE DIFFERENCE of the top-10 features

4. ROOT MEAN SQUARE PAIRWISE DISTANCE of the top-10 features. Then plot the redundancies you obtain using method 1 versus the usual metrics of sensitivity, specificity.... Do this again for method 2 and method 3. See if any of these results fit linear/logarithmic/exponential...any kind of correlational fit. Even if no correlation is obtained, this data will be very useful for future investigators.

5. Scope: The title promises a broad biomarker review and an AI comparison; the review concentrates mainly on A β and tau with brief mention of alternatives, omitting other established biomarkers (e.g., NfL, neuroinflammation, structural/functional imaging breadth). Hence, the title and associated verbiage in the manuscript may need to be changed to "A Review of Alzheimer's Disease Biomarkers A β and tau and a Comparison of Predictive Artificial Intelligence Methods for Diagnosis"

6. Factual errors: Ridge regression is described as using fewer variables, which is incorrect (ridge shrinks but typically does not select variables). Choice is not specified. Minor methodological misstatements (e.g., generic 'gradient descent' for boosting) also appear. The 'Redundancy' metric is not defined. Please define what exactly is meant by 'redundancy' both in mathematical and conceptual terms.

7. Assumptions: Assumes behavioral variables sufficiently proxy A β /tau without direct measures; assumes XGBoost superiority largely from nonlinearity; does not justify model hyperparameters, class balance, or kernel choices. Add justifications for assumptions, hyperparameters, class balance and kernel choices.

8. Consistency: Text alternates between 'three' and 'five' models; some explanatory claims conflict with standard properties of algorithms

9. Robustness: Single synthetic dataset with no external validation, sensitivity analyses, or seed/parameter robustness checks; no ablation or preprocessing. Perform seed/parameter robustness checks and explain how preprocessing was done, perform ablation studies; performing ablation studies is important because you are discussing redundancy.

10. Statistical analysis: "Comparisons use t-values/p-values without specifying the sampling unit (folds/runs) or paired design; no confidence intervals, effect sizes, or multiple-comparison correction. ROC-AUC, PR-AUC, calibration, and McNemar/DeLong tests are absent.

11. Corrections: no explicit adjustment for confounders (e.g., age, sex, education) or sampling bias; no propensity weighting or stratified analyses. Justify why this was not done and put down in limitations.

12. Range: Feature distributions and coverage are not shown; no analysis of performance across subgroups or extremes (e.g., very old, severe comorbidities). Justify why this was not performed or put this down in limitations.

13. Collinearity: Spearman heatmaps address pairwise correlations qualitatively, but no VIF, PCA, or regularization diagnostics are provided; redundancy metric lacks a formal definition or thresholding.

14. Normalization: Feature scaling/normalization is not described. For SVM and KNN, standardization or min-max scaling should be specified and applied within cross-validation folds to avoid leakage.

15. Experimental design: Software libraries are listed, but train/validation/test strategy, number of runs/seeds, hyperparameter search, class imbalance handling, and code/data release are not detailed. Nested cross-validation, hold-out external validation, and pre-registration would improve rigor. This manuscript presents AI research from an interesting new angle. However, the concerns above must be addressed in order for this contribution to attain its full potential. I urge the author to address these concerns. I look forward to a revised manuscript.

I had a quick question, but is it okay for me to use a synthetic dataset or would you prefer it being replaced with real data? Or, I can keep the synthetic one and use the real data as external validation, because I know that's one of the comments you gave.

Dear author,

If you are using a synthetic dataset, you will need to validate it against a real dataset. If you are unsure of how to do this, input this prompt into chatgpt "how to validate synthetic dataset against real dataset"

Please also address other review comments and see if you can make comment 1 work.

Please let me know if you have any questions or need a video call.

Best,

Shireesh Apte

Thank you so very much again for taking your time talking to me last Wednesday. I appreciate it very much. I have made quite a few updates to the paper since then. My paper is now framed as a cross sectional diagnosis classification study with more careful discussion of class imbalance, balanced accuracy, calibration, and the sensitivity/specificity tradeoff. I also explicitly documented gridsearchCV tuning, added balanced accuracy to the reported canonical metrics.

I clarified the fixed classification threshold and the synthetic-noise experiment's limited purpose. I worked throughout the past weekend. It just that everytime I read it, I thought of something else to add/change, that's why it took so long. My apology for the delay! I still need to complete table 1 and references; give it a new title and polish my writing, but I have all the major updates in.

Attached please find a revised draft copy. I will continue updating my paper throughout this week, and submit the final version this weekend. Thank you again and I would appreciate your comments and thoughts.

How long will it take you to run feature importance 1-10, 11-20, 21-30 and 31-40 for AD discrimination for all ML models? (see attached chatgpt pdf file).

Shireesh Apte

Thank you so much for getting back to me so quickly and for the thoughtful suggestion.

I understand why comparing multiple alternative subsets, like rank-based or domain-specific groups, could be interesting. But I think that would be testing a somewhat different question from the one my current paper is focused on.

In my paper, I was trying to study whether performance stays relatively stable after reducing overlapping predictors to a smaller representative feature set. I was not trying to claim that many different arbitrary subsets would all perform equally well. I think the broader subset-replacement experiment you suggested could be valuable, but it would lead the paper in a different direction and

support a broader claim than the one I am making now. If I go that route, I would probably have to redesign a big part of the revision, rerun all the analysis and rewrite much of the paper.

I would definitely be interested in exploring that idea in the future. For this revision, though, would it be okay for me to keep the paper focused on the narrower claim, so that the manuscript stays aligned with the experiments I actually performed?

I also wanted to be honest that rerunning and incorporating a broader new experiment would take me a lot of additional time. As you know, the first revision already took me a long time to complete. I spent a lot of time running analyses, checking the results carefully, and then revising the paper so the claims matched what I actually tested. I worked on it over many weekends and late nights, and I was able to use my holidays, winter and spring breaks to make progress. At this point, though, I also have AP exams, schoolwork, and college applications coming up. I am concerned that if I run a new analysis in a different direction and rewrite the paper to reflect the change, I may not be able to bring the paper to a finish in a reasonable timeframe.

I would really appreciate your understanding and consideration, and I'm very open to clarifying the paper's wording further so that the claims stay careful and accurate.

Thank you again for your time and guidance.

Understood. However, your study carries insufficient novelty - it is a re-framing of existing methods.

What would you say if you added a 'Perspectives' section where you include this idea as well as the redundancy ideas we discussed earlier? This way, you will not have to perform any new experiments.

If you agree, I will write the content and send to you for approval.

Best,

Shireesh Apte

Thank you very much — I really appreciate that suggestion. Since I am still learning a lot about academic writing, I honestly am not even familiar with what a perspective section would look like in this context. But the idea makes sense to me as a way to acknowledge the broader direction without changing the main paper into something different from the research I conducted.

I think adding a Perspective section could be very helpful. I also appreciate your willingness to write the content, especially since that broader subset and the redundancy regression pattern ideas are quite interesting although they go beyond the exact question I tested in my current paper.

Thank you again for your guidance and for helping me think through this so carefully. I really appreciate it.

please incorporate the attached “perspectives” section into the manuscript as discussed. You can drop it into the appropriate place in the manuscript (probably after the discussion section).

Also change the abstract and conclusion to include a line or two from the perspectives section.

Thank you for addressing my comments. Accepted.