**Peer Review and Communication History**

**MS Title**: Machine Learning Mega-Analysis Applied to the Response Time Concealed Information Test: No Evidence for Advantage of Model-Based Predictors Over Baseline

**Author Names**: Gáspár Lukács and David Steyrl

**Submitted:** July 15, 2021

**Editor First Decision**: Revise & Resubmit

Jan 13, 2022

Dear Gáspár Lukács,

I have now received two reviews of your manuscript, “Machine Learning Mega-Analysis Applied to the Response Time Concealed Information Test: No Evidence for Advantage of Model-Based Predictors Over Baseline”. I thank the reviewers for taking the time to provide their expert opinions on the manuscript, particularly during such a difficult time, and I apologise for the delay with the decision letter. The reviewers had largely positive reactions to your manuscript, though they raised some points that I would like to see addressed in a revision. I also read the manuscript both before and after receiving the reviews, in order to form my own independent opinion and then consider the manuscript with the additional context of the reviews. While I’m not an expert in ML or RT-CIT, my perspective was mostly positive, and like Reviewer 1 the analyses seemed robust to the best of my understanding. However, like Reviewer 1 I wondered how the inclusion of error RTs or cognitive constructs via RT models might change performance, and like Reviewer 2 I thought that the clarity regarding the motivation, purpose, and expectations of the project were somewhat unclear. While I do not have anything to add beyond the reviewers’ points, I think these points are important ones, and should be addressed in a revision. Specifically, I would like to see the points brought up by Reviewer 1 at least given some space in the discussion to help guide future research on the topic, and I would like to see the remaining context requested by Reviewer 2 incorporated into the appropriate sections of the manuscript (mostly early on, I think). Therefore, I strongly encourage you to submit a revised version of the manuscript for further consideration at Collabra: Psychology. The revision should include a document with a point-by-point response to the reviewers’ comments, outlining each change made in your manuscript or providing a suitable rebuttal.

Overall, I believe that your manuscript is of excellent quality, and I hope that you will choose to revise it for further consideration at Collabra: Psychology. I look forward to receiving your revision.

Please ensure that your revised files adhere to our author guidelines, and that the files are fully copyedited/proofed prior to upload. Please also ensure that all copyright permissions have been obtained. This is the last opportunity for major editing, therefore please fully check your file prior to re-submission.

If you have any questions or difficulties during this process, please contact us at the editorial office [editorialoffice@collabra.org](mailto:editorialoffice@collabra.org). If you have any further questions about the reviews or revisions, then please feel free to contact me at [nathan.j.evans@uon.edu.au](mailto:nathan.j.evans@uon.edu.au).

We hope you can submit your revision within the next six weeks. If you cannot make this deadline, please let us know as early as possible.

Sincerely,

Nathan Evans

# Reviewer 1

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The paper presents an interesting machine learning analysis of results from the RT-CIT - a paradigm used to understand whether a person is concealing knowledge (for example in studies of criminals who know what the murder weapon is). Typically, data from the RT-CIT has used the difference in RTs between probe and irrelevant items, however this study proposes adding additional variables to understand whether the classification accuracy can be improved, and the authors test this through a machine learning “mega-analysis”. The results showed that classification accuracy was not improved significantly over baseline when adding additional predictors. I had not previously read about the RT-CIT task, however have attempted to acquaint myself with the literature briefly.

The study and methods appear well conducted (and thorough in the cross-validation loops), and I commend the authors both on the openness of the methods/results and on publishing, in effect, a null result. The analysis is clearly set out and thoroughly performed, and I can envisage similar ML procedures being used for other questions of interest for similar studies. I would recommend this article for publication following some very minor comments being addressed.

First, an accuracy of around 70% seems quite low for accuracy, this means just under a third of participants are misclassified. Can the authors comment on this and the potential repercussions of the low accuracy. Is this an artefact of the ML method or the task? And if it is the task, could the authors comment on whether adjustments should be made to the standard method/analysis to improve this accuracy?

Secondly, the authors state that RTs are taken from correct only trials, which is reasonable. However, could error RTs, or further, modelling the data with an evidence accumulation framework, provide any further information to improve classification? The masters thesis by Strahm (2017) seems to suggest this is the case, and this is noted in the discussion. Although beyond the scope of the current study (the modelling at least), I feel more weight should be given to the investigation of a process model, especially when motivating the task for future use.

Finally, was random forest considered as a classifier in addition to the three methods used? And if so, was there any particular reason not to include it (I.e. speed or not increase to performance)? Random forest tends to perform better when there are a high number of noisy features, which may be the case with RT data, so it surprised me that it was not included.

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

# Reviewer 2

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Overall, I find the reporting of the methods and analyses to be thorough and transparent. I commend the authors for registering their analyses, specifying deviations from the registration in the manuscript, and providing thorough supplemental materials for further scrutiny. My appraisal of this paper is generally positive. My criticisms concern the conceptualization of the project and matters of clarification.

Unfortunately, I am not an expert on ML, so I cannot comment in detail about the analyses the authors performed. I hope that another reviewer is able to provide more informed comments on the analyses than I am.

My central concern is that the purpose of the project and the researchers’ expectations are somewhat unclear to me. The manuscript opens with a preview of the results: that this is a case where ML does not seem to offer meaningful improvement over simpler approaches. But it isn’t clear to me from the manuscript or from the preregistration whether the authors expected this to be the case. The author’s expectations do not bear directly on the appropriateness or informativeness of the analyses themselves, but they do help contextualize the process leading to these tests. That is, if the authors expected these additional features to offer little to no advantages over the more conventional approach, that expectation may have shaped the decisions about what features to include and the plan of what analyses to conduct.

Relatedly, I would like to know more about the situation at the time of registration and planning of these tests. This project makes use of data that the authors were already familiar with, so their previous knowledge of the data and prior analyses of the data are relevant for contextualizing the meaningfulness of the analyses reported here. If the authors already had knowledge about the relationship of these additional features tested in these ML models to the outcome variable, that knowledge would reduce the informativeness of these analyses. In the worst case, the present analyses might represent a computationally intensive recapitulation of what the authors already knew about these data, and cross validation of the models would offer no protection for this issue. Along these lines, researchers who have worked on registration templates and other open science issues relating to secondary data analysis have emphasized the importance of prior knowledge of the data (see, e.g., van den Akker et al, 2021, <https://psyarxiv.com/hvfmr>). To be clear, I am not suggesting that the authors are attempting to conceal anything (quite the contrary, the work is admirably transparent). Rather, I want to make extra sure that this these tests are as credible as they seem to be.

Recommendation

As I said above, I have a generally positive appraisal of this paper. I would like to see further clarification on the authors’ prior knowledge of the data and their expectations/hypotheses, since I think these are highly important for contextualizing the informational value of this project. To the extent that the authors’ prior knowledge of the data influences the credibility of the conclusions, the authors should comment on this issue in a revision.

I request that the authors add a statement to the paper confirming whether, for all studies, they have reported all measures, conditions, data exclusions, and how they determined their sample sizes. The authors should, of course, add any additional text to ensure the statement is accurate. I also request that the authors add statements describing any other related studies, including pilot tests, they have not reported here. This is a modified version of the standard reviewer disclosure request endorsed by the Center for Open Science [see <http://osf.io/hadz3>]. I believe the authors have already lived up to this request, but I include a version of this statement in every review.

I always sign my reviews.

Timothy J. Luke

Minor Issues

The authors note in the methods section and preregistration that they excluded data from participants with an accuracy rate lower than 75%. Some brief additional clarification on why this cutoff was used might be helpful.

In the General Discussion, on line 376, the authors state “The correct way to [incorporate additional characteristics] is via ML-based multivariable models and proper cross-validation.” This strikes me as an overstatement. Surely this is not “the” single only correct way to attempt this.

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

**Author Response**  
Jan 24, 2022

See Supplement “v2\_1555879-replies-to-the-reviewers-comments”

**Editor Final Decision:** Accept

Feb 16, 2022

Dear Gáspár Lukács,

Both of the previous two reviewers and I have now had a chance to read over your revised manuscript “Machine Learning Mega-Analysis Applied to the Response Time Concealed Information Test: No Evidence for Advantage of Model-Based Predictors Over Baseline”, along with the letter describing the changes you made. The reviewers were both extremely satisfied with the changes that you made, and I am happy to say that your paper is now officially accepted for publication in Collabra: Psychology. Congratulations on this excellent work, I think it will make an important contribution to the literature and I look forward to seeing it published! I hope your experiences with Collabra: Psychology have been positive and that you will continue to consider it as an outlet for your work.

As there are no further reviewer revisions to make, you do not have to complete any tasks at this point.

You will be receiving separate correspondence regarding any production and technical comments, data deposits, as well as publication charges. We work with the Copyright Clearance Center to process any applicable APC charges. Please note that your APC transaction must be completed before your article gets published.

You will have an opportunity to check the page proofs before we publish your article. Thank you again for publishing in Collabra: Psychology.

Sincerely,  
Nathan Evans

# Reviewer 1

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The authors have thoroughly addressed all of my major concerns and I am happy for this manuscript to be published in the current form.

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  |  | ✔ |  |

# Reviewer 2

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The authors have done a thorough job responding to my previous comments, and I believe the revisions they’ve made to the manuscript address my concerns adequately.

A point of clarification about my concern about the authors’ expectations – in their response, the authors say that there wouldn’t have been much of a point in doing the project if they did not expect any improvement. I appreciate that point, but I think I disagree. One might do the project expecting no improvement for exactly the reason the authors make: that successes with ML are widely publicized and failures are inadequately recognized. That is, one might try something they don’t expect to work exactly for the purpose of demonstrating that it doesn’t work. That said, it would be an awful lot of code to write and a lot of computational power to muster just for the exercise of demonstrating that something doesn’t work.

In any case, I commend the authors on their solid work.

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  |  | ✔ |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  | ✔ |  |  |