Reviewer: 1  
  
Comments to the Author  
The broad range of breath-to-breath data collection and treatment options combines with different exercise protocols often presents a confusing picture and interpretation to exercise testing.  Thus, this scoping review has an important imperative that, in certain respects, speaks to addressing reporting concerns regarding the rigor and reproducibility of data processing, analysis and interpretation.  The authors are to be congratulated for the breadth of their reach and employment of machine learning and other state-of-the art methods.  Against this enthusiasm are the following concerns which temper enthusiasm:  
1. Where is the physiology? Why do these procedures matter?  Where is the convincing evidence that errors in reporting, interpretation and physiological understanding have resulted from the omissions etc. identified herein? Without convincing the reader of such any guidelines developed on the basis of the current findings have the danger of being perceived as meaningless over-reach.  
2. One of the largest sources of error in exercise testing regards the VO2peak versus VO2max issue that, especially in patient populations, has raised substantial concerns especially of late.  The present article avoids the issue of validation tests which can be used to construct the classical VO2max plateau even in the absence of a levelling off in the CPET VO2 (see Poole et al. Eur J Appl Physiol. 2008 Mar;102(4):403-10 and Poole & Jones, J Appl Physiol (1985). 2017 Apr 1;122(4):997-1002). This is a crucial issue that some of the present findings – with appropriate and closer individual article scrutiny and expertise – might address.  
3. Methods 2.1. Doesn’t the validation step remove many of the concerns, at least for VO2max?  
4. Page 2, line 55. Typo. “paywall”  
5. Page 4. Should probably complete individual words from the quotations.  
6. Page 5, lines 29-35. What is important about the specific RER.  Some physiological consideration necessary.  
7. Page 5, line 51. And what exactly is the “anaerobic” threshold? Please inject some physiology here.  
8. Page 8. Lines 51-3. Show numerical examples.  How does this impact interpretation and physiological conclusions based thereon?  
9. Page 9.  “30 s average” So, is this good?  How does it fit with physiology and clinical assessment?