0.1 General Comments

We very much appreciate the opportunity to submit a revised version of our manuscript and are grateful for the very constructive comments of the reviewer. Overall, the manuscript has undergone a major revision to comply with the reviewer's suggestions. The major changes are outlined below, together with detailed specific comments relating to the points raised by the reviewer.

In the previous version, we introduced a co-FTF algorithm that was shown to be competitive and useful for data analysis with measurements generated from multiple views. The reviewer pointed out two key shortcomings of the previous work: (i) the paper did not properly address the interests of the chemometrics community in both presentation and data analysis, and (ii) the novelty and clarity of the work was not appropriately presented. The paper was carefully reconsidered to address these shortcomings.

In this revision, write-up of the algorithm was significantly simplified and the experiments illustrating the co-FTF were completely reexamined. A comparison to data fusion with PLS and extensive analysis on the separate biology and chemistry cases were added in the experiments section. New experiments to address robustness to a permuted response and variable importance were considered, and the advantages of co-FTF were presented. In addition, execution time of the algorithm was added, considerations for co-FTF parameter estimation are now discussed, and the data set used for analysis is now available (in blinded format) in an **R** package that demonstrates the co-FTF algorithm (included as supplementary material).

The revised work focuses on the important contributions of the proposed approach in both the classification and data analysis settings, rather than on extensive benchmark comparisons with other multi-view algorithms. We believe this change in focus enhances the novelty of the work, by showing that co-FTF can perform variable importance, is robust to a partially permuted response, and obtains performance gains over other techniques with similar objectives. The extensive numerical results in the previous version established the overall competitiveness of our approach, and we would be happy to include those if the reviewer considers it necessary.

Next, we address the specific comments of the reviewer:

0.1.1 Reviewer One Response

Comments to the Author This manuscript reports an approach that allows training classifiers on multiple data sets (the authors call these views) with the aim of getting classification performance from the multiple set that surpasses that from the individual sets.

This article is formatted and presented in a style that is not well-matched to this journal. It also has some overlap with other work from these authors that is being published or has been published elsewhere in a computational statistics journal, but (so far as I can tell from imperfect information) the emphasis there is on the algorithms. Nevertheless, I wondered about the real novelty of work presented here. If the authors address my criticisms below, that concern will evaporate, I believe.

The idea presented here is not easy to grasp, given that it is presented in what amounts to a foreign language, but the concept is akin to data fusion something that the authors dont mention in a direct way, if they mention it at all.

From the advice of the reviewer we now compare the proposed approach to data fusion with partial least squares. We found this approach to be both insightful and interesting, and thank the reviewer for pointing out the connection to data fusion.

However, it is an interesting idea, and it may well be related to work already done with bagged, tree-based classifiers and/or Bayesian nets. The authors are not very generous with citations to other work in data analysis, and I think that they should offer a bit more in the way of an introduction to put their work in clearer perspective. That means discussing alternative approaches a little and referencing other work.

Several references from data analysis were added to address this point including [4,6,7,18] among others. The analysis measurements are now more similar to the ones considered in these papers, such as sensitivity and specificity, as opposed to kappa.

Second, the methodology is applied to a very poorly specified data set. If the set is proprietary, fine, but at least explain how many objects we have and how many features. Their explanation simply does not add up : we have a 438x431 data set made from 151 binary chemical predictors + 191 continuous biological descriptors ¡¿ 438 or 431, so we dont seem to have the full story on the data. If the data are not proprietary, we need a good deal more explanation, as well as access to the data. And we need to know more about the descriptors. We need to know what the class information is what is an adverse event (AE) anyway?

Upon rereading the paper with respect to this point, we agree that the data discussion was lacking in the previous version. The data discussion is more thorough in this revision and a blinded version is available for analysis.

Third, the work has no real punch line. We fail to see if/how this method is superior on this classification to other, established classifiers and their competitors. We see (no surprise, really) that a random forest outperforms the SVM here, that the biology contributes just enough to help get us past limit set (by the authors?) on the seemingly arbitrary kappa parameter, etc, but we dont see how established classifiers do on each set, any why we should care to read more about the fused classification (sorry, the co-trained algorithm for multi-view data). So, what happens if we randomly permute classes, as is commonly done for classifiers: does this fall apart? How long does this analysis take? Is there really a benefit to this approach, give the time difference?

In the revised version, we provide an insightful comparison to PLS and data fusion with PLS, which we believe provides a real comparison. A separate and combined analysis is now performed. The kappa parameter was replaced with sensitivity to address the above point. The suggested experiments were added. Overall, we believe the comparisons performed in the revised manuscript address the above comments by highlighting the benefits of the proposed approach in both classification and data analysis.

And, can this be demonstrated on a simple data set that allows testing by others?

As noted above and in the paper, we prepared an **R** package with the data set included for this purpose.

Fourth, the authors need to take some care in revising to follow the norms and language of the journal. The references are incomplete, the referencing style is non-standard (please have a look at the journal), and the figures dont meet requirements. The authors dont even have the corresponding author indicated, or the institutions, and no key words.

All of the above points were addressed in the revision. In addition, the specific criticisms listed at the end of the review were each addressed in this revision.