REPLIES TO REPORT 1

1. *Section 1, page 1, paragraph 2. "... sets indirectly constraints to the New Physics ... " should read "... sets indirect constraints on the New Physics ...". In the same paragraph, I find "used for their discovery" a bit vague. Rather the authors could write that the coupling to photons allows for *detection* of ALPs.*

   Fixed

2. *Section 4.1, page 4, paragraph 3. Is this efficiency with respect to all 3 mass points combined or an average efficiency for the 3 mass points - please clarify this point.*

   This refers to the efficiency for the ALPs with the 3 different masses. We've made this explicit in the text.

3. *Figure 1. The figures are rather small here. I would ask the authors to increase the label sizes for better visibility. It is also more appropriate in the caption to refer to the figures with (left), (centre) and (right) or at least use the same labelling in the caption as the figures (eg "No conversions vs 0CV NN")*

   We have improved the plots, this time adding error bars and reducing the binning.

4. *Section 4.2 - See main report for discussion of choices of the training parameters.* This is a fair point, which was also raised by the other referee. Our choice of features was based in the obtention of maximum discrimination. Moreover, we use the outputs of two other classifiers as input to ours. Adding the full list of features to a single classifiers and retraining it for our specific topology could also have improved the performance, and could be done as a refinement in the future. However, the two classifiers whose output we use as inputs to ours are highly complex, as explained in Ref. [27]. Doing this would have required a very significant amount of work and we were really interested in deploying classifier in time for the 2018 data taking. So that is why we decided to proceed as we did. We have made some corrections to the text to clarify all these points. This can be found in the paragraphs before and after the list of features.

5. *Section 4.2, page 5, paragraph 2 - Add statistical test for comparison of training and test samples (see main report)*

   We have added the result of a Kolmogorv-Smirnov test (see new Table 4).

6. *Section 4.3 - Add discussion of the performance in terms of processing time and whether an optimisation in this respect was studied (see main report)*

   In Ref. [25] two examples can be found showing how the NNDrone framework allows to reduce the processing time by an order of magnitude while keeping the same classifier performance. As stated in the paper, the structure of the classifier was kept simple in order to allow for quick evaluation. We have added a small discussion about this.

7. *Table 4. The MLP requirement is presumably a lower bound on the "classiffier" quantity from Figure 1? Please clarify this in the table or in the caption. While the expert reader can probably assume this, the general reader might not.*

   Fixed

8. *Section 5, page 6, paragraph 3 - See main report for adding details on the sensitivity estimate 9.*

   We agree having more details could have been useful. However, since this is catalogued by the collaboration as a "Detector performance" paper, the amount of information about the physics sensitivity we can provide is limited. For the same reason, the sensitivity study we performed does not include a full optmization of the search and we simply make hypotheses of the discrimination that we could achieve.

9. *Appendix A. Not clear what "combination" of all signal models. → see request #2.*

   Added same clarification as above

10. *In section 2, page 2, paragraphs 1 and 2 describe the event generation for the B_s^0 and ALP signals which are different for the two. It would have been useful to know if the trigger performance would be sensitive to the choice of generators and whether other generators were investigated, or whether such an investigation would not be possible.*

    We took this paragraph from the "Report" section assuming it was also a requested change or question. For B meson decays, LHCb uses a specific configuration (including PYTHIA and EVTGEN) that has been tuned specifically for the experiment and that it is known to work reasonably well (most of LHCb's flavor physics programme is based on this configuration). For ALPs, for the model described on Ref. [20], we had really no alternative configuration. In any case, MadGraph works at NLO level and is widely used as a reliable source of MC simulation across different experiments. For the actual search for and ALP or $B^0 \to \gamma\gamma$ we agree this comparison of efficiency for different generators might become important. However, in this case we were also limited by the need to have this trigger lines running in time for the 2018 data taking, which in some cases prevented more detailed studies.