REPLIES TO REPORT 2

1. *why at the beginning of the introduction Bs and not Bd is mentioned?*

   This is now corrected.

2. *While I agree that the QCD axion would couple to the gluon, is this a generic feature of anything that people would call ALP?*

   This is now clarified: "For the type of model addressed in Ref. [7], ALPs couple... "

3. *It is not clear to me which L0 selection is implied here and which effects it has on the downstream analysis. This is barely mentioned at some point, but more details should be given.*

   We agree this was not completely properly described. We have now improved the description to make it more clear. Furthermore, the selection that was previously described at the end of section 2 is really only used for the training of the classifier samples, but not accounted for in the HLT1 efficiencies in Table 2. We have also corrected that.

4. *The choice of giving a cryptic code snippet to describe the network architecture is quite low level and dumps on the reader the task of retrieving the meaning of any parameter from the scikit learn manual. Under the assumption that the results don't depend on the library used for the implementation (I don't doubt this), it is preferable to have a text-based description of the architecture, the hyper-parameter setting, the training procedure, etc.*

   A more description of the architecture, together with the motivations for the choice of parameters, is provided now, before the code snippet.

5. *the choice of input features is very arbitrary and leaves a lot of open questions. Given the demonstrated ability of Deep Neural Networks as feature extractors, what is the point of using as input the score of another ML algorithm instead of the inputs used to train those algorithms?*

   This is a fair point. Our choice of features was based in the obtention of maximum discrimination. Adding the full list of features to a single classifiers and retraining it for our specific topology could indeed improve 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 decided to proceed as we did. We have made some corrections to the text to clarify these points. Before the list of features, we have added "We select those providing significant discrimination between signal and background.". Then, after this list, we've added a small discussion about the inputs that are outputs of other classifiers. "It should be remarked that, as mentioned above, our classifiers use as inputs the outputs of two other different classifiers, one designed generically to identify calorimeter photons (γ shower shape) and one to differentiate photons and π^0 mesons (γ prob). While, ideally one could integrate the features used in these classifiers into our own one, making use of the ability of algorithms such as Deep Neural Networks as feature extractors, it was decided against this given the high level of complexity in the training of both γ shower shape and γ prob. The first accounts for the expected deposit shapes in different regions of the calorimeter and excludes the possibility that these are originated by a charged track. The second also relies on the energy value and the calorimeter zone comparing the expected energy deposit of a photon with respect to that of a π^0. Although both classifiers were trained to maximize its sensitivity for B meson decays, potential extensions of our work could involve integrating all the features into a single classifier to be trained specifically with our signal samples.". We have also added a reference to [27], where these classifiers are described.

6. *Fig.1 should be remade using histogram line rather than dots. In these plots and in those shown in appendix, one would need to use more statistics and/or less bins to reduce the effect of random statistical fluctuations*

   We have improved the quality of the plots in the main text, adding error bars and reducing the binning. Concerning those of the appendix, their edition is in control of one of the authors, and he has unfortunately left the field. so changing them would mean a significant delay for publication. Since they're only part of the appendix, we'd kindly ask you to accept them in their current form.

7. *In Section 4.4, I would have expected some discussion about the latency, memory footprint, and in general some quantification of the performance.*

   The other referee also stated their concern about this issue. In Ref. [25], two examples can be found in which 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 sentence.

8. *A plot should be added to show what is said at the end of Section 5*

   Such a plot would essentially look like Fig. 1 in Ref. [7], so we prefer not to.

A few editorial remarks

1. *In the abstracts, ALP is not a b-decay. An ALP-mediated decay to a diphoton final state would be a decay*

   Fixed

2. *In the abstract, The fast → A fast*

   Fixed

3. *In the introduction (first sentence) final state *of a B meson**

   This "final state" sentence is trying to refer both to ALPs and B, so we think it is fine as it is.

4. *In the introduction, Standard Model of particle physics*

   Fixed

5. *In the introduction, L0 is undefined*

   Fixed

6. *In section 4.1, data is plural*

   Fixed

7. *In Table 4, the first column is not the efficiency. It is the process. The other three columns refer to the efficiency.*

   Fixed

8. *In section 5, Sect. → Section*

   Fixed