Report

The authors investigate anomalous quartic gauge couplings (aQGCs) at a high-energy muon collider in the process $\mu^+\mu^- \to \nu\bar{\nu}\gamma\gamma$, focusing on dimension-8 SMEFT operators. They introduce the Nested Local Outlier Factor (NLOF) as a novel anomaly detection method and demonstrate, using `MadGraph` and `Delphes` simulations at $\sqrt{s} = 3$ and 10 TeV, that NLOF significantly outperforms standard LOF, k-means anomaly detection (KMAD/QKMAD), and cut-based strategies. In particular, NLOF yields cross sections with a much stronger dependence on the Wilson coefficients after anomaly-based event selection, leading to bounds up to an order of magnitude tighter than with LOF, KMAD/QKMAD, or cut-based methods. The study shows that for several operators, especially of the T-type, the muon collider can probe effective scales well beyond the LHC reach, highlighting both the physics potential of muon colliders and the usefulness of advanced unsupervised learning methods for EFT searches.

The paper overemphasizes the "unsupervised" nature of the method and lacks comparison with key baseline algorithms. These issues should be addressed as outlined below. I thus recommend a **major revision**. Provided the authors implement these revisions, the paper has the potential to make a valuable contribution and should be suitable for publication in *SciPost Physics*.

1. The method is presented as fully unsupervised, yet in Sec. 4 the thresholds a_{th}, Δa_{th} and the choice of k are optimized using signal significance. This introduces a degree of supervision, effectively tailoring the method to the specific signal model under study (dimension-8 aQGCs). True unsupervised anomaly detection should define thresholds from background data alone. I strongly recommend that the authors clarify this point, and ideally include results where thresholds are set purely from background distributions and signal is used only for evaluation.

2. NLOF is compared against LOF, k-means anomaly detection (KMAD/QKMAD), and cut-based selections. These are not the most relevant state-of-the-art baselines. Since NLOF is a density-based method, more appropriate comparisons would include DBSCAN, OPTICS, Isolation Forest, kNN-based distances, etc ... Even a limited comparison on a subset of these would provide a much more convincing demonstration that NLOF offers genuine improvements.

3. The current results focus only on cross sections after optimized cuts and the resulting EFT coefficient bounds. It would be very valuable to also show ROC curves and AUC values for LOF, NLOF, k-means, and cut-based selections. These are threshold-independent and directly illustrate the separation power of the anomaly scores. This would also mitigate the concern about "cheating" via signal-based cut optimization.

4. In Sec. 4 the anomaly score thresholds are tuned to maximize significance. This mixes background rejection and signal efficiency. To make the comparison more

transparent, I recommend showing significance values at fixed background acceptance levels (e.g. 1%, 5%, 10%). This would highlight the gain in signal efficiency provided by NLOF relative to LOF or k-means, independent of cut tuning.

5. A natural and widely used extension is to train an autoencoder on SM data and apply anomaly detection (eg. LOF, NLOF) in the learned latent space. This often improves robustness and separation power. Even a short discussion of this possibility would broaden the impact of the work.

6. Figures 2–4 show cross sections and coefficient bounds after optimized cuts. These effectively fold in the anomaly detection performance, but they do not directly show how well LOF/NLOF separate signal from background. Adding anomaly score distributions and ROC curves alongside these figures would make the connection clearer.

7. Finally it would be helpful to state the computational cost of NLOF relative to LOF.