Response to referees - An Autoencoder-based Online Data Quality Monitoring for CMS ECAL

Dear editor,
We would like to thank the referees for reviewing this paper and furnishing this report.
In general, a few details were omitted in the first version due to the constraint of 6 pages for the proceedings report.

However we have carefully considered all comments, and have applied the necessary changes to the original version of the paper to address the issues raised. Detailed responses to all the comments can be found below.

Sincerely,
The authors

Color code:

BLACK: question from the referee.

BLUE: answer to the question.

GREEN: new text in the paper.

Reviewer 1

1. additional clarity on how the model was tested and validated, especially in comparison to traditional methods, could strengthen its claims
A: We have added a comparison with a baseline study based on cut-based traditional anomaly detection algorithm in the text in section 4.1 and added the results in Table 1.
Section 4.1

2. more information on the autoencoder architecture, training details, and computational requirements could be beneficial for replicability and understanding the technical complexities involved

A: We have added a figure for the model architecture and some details of the network in the text in Section 3.1

The architecture of the AE used is shown in Figure 2. Each "ResBlock" consists of two

convolutional layers, with a ReLU activation [12] and a residual mapping. The input image is passed through the encoder network that consists of a CNN, followed by an aggregating maxpool layer.It is then sent through sequential layers of ResBlocks, up-sampling the feature maps progressively.
This is followed by a global maxpool that creates a compressed dense layer of the encoded space. The encoded layer is then passed to the decoder network as the input, which performs these operations in reverse and outputs a reconstructed image. Three separate models are trained with this architecture: one for the barrel and one for each of the two endcaps, owing to the differences in their shape, granularity, and response.

Reviewer 2

1. **Could you mention the specific architecture of the model? Which ResNet?**
   A: We have added a figure for the model architecture and some details of the network in the text in section 3.1 (Please refer to answer to question 2 of reviewer 1)

2. **I see that spatial correction improves the results for endcaps. But I was wondering why you need it. Shouldn't the model learn it? If possible, please elaborate on this.**

   A: The effect of the spatial correction is related to the occupancy pattern across the detector. The detector has higher occupancy as it gets closer to the beam pipe, and this occupancy "gradient" is more prominent for the endcaps than the barrel as mentioned in the text.
   As you point out, indeed the model learns this occupancy gradient, and precisely because of which it gives different losses for the same anomaly on different parts of the detector.
   So a zero-occupancy tower on the outer edge of the endcap, where the nominal occupancy is lower, will get a smaller loss value while a zero-occupancy tower on the inner edge, where the nominal occupancy tends to be higher, will get a higher loss value, which will bias the anomalous loss threshold, leading to a lot of false detection.
   By normalizing the loss with the average occupancy map, the zero-occupancy anomaly loss is "flattened out" and the FDR improves a lot.

3. **Could you elaborate on why spatial correction leads to worse results for hot tower anomaly?**

   A: Spatial correction is introduced to "flatten out" the pattern across the detector for the anomalous loss (see the answer for the previous question). However, this only works for the zero occupancy anomaly scenario where the anomalous loss will be proportional to the nominal occupancy of the tower, as the model reconstructs the nominal occupancy instead of the zero occupancy.

In the case of the hot tower anomaly, the model will still reconstruct the nominal occupancy of the tower, so the loss will be higher for the regions with lower nominal occupancy. This makes the hot tower anomalous loss show the gradient that is "opposite" to the zero-occupancy case. If we use the "reverse" occupancy map for the spatial correction, hot tower anomaly results will improve and zero-occupancy tower results will worsen from the spatial correction. We decided to keep the spatial correction using average occupancy map such that it favors the zero-occupancy tower scenario, because zero-occupancy towers are generally harder to detect than the hot tower case and the FDRs for the hot tower case that are worsened from the spatial correction can be improved with the time correction.

To address questions 2 and 3, we have added the following sentences in the text in Section 4.2:
Spatial correction improves the AE performance for the towers with zero occupancy anomalies. Without the correction, the loss values for the zero occupancy towers are proportional to their nominal occupancy, which is biased to be larger in the higher $|\eta|$ region.

This gradient in loss is flattened out by the spatial correction, and the effect of which is more pronounced for the endcaps due to their larger effective gradient in the occupancy. For hot towers, spatial correction in turn increases the FDR since the hot tower loss is biased to be higher in the opposite direction, towards the lower $|\eta|$ region.
So while the spatial correction reduces the gradient in loss for zero occupancy towers and improves their detection, for the hot towers the gradient is enhanced and the AE performance slightly worsens. However, this effect is mitigated by the time correction that gives an order of magnitude improvement in all anomaly scenarios.

4. **Can you briefly mention why the zero-occupancy towers in Fig 5b are not detected in Fig 5a?**

   A: The zero-occupancy towers in Fig 5b (Fig 6b in the updated version) are the towers that are known to have had issues from the previous data-taking era. The autoencoder model has learned the pattern of these "dead towers" from the training data. During data-taking, these towers with known issues are masked and are not taken into account for the inference of the model, and thus they do not show up as red in Fig 5a (Fig 6a in the updated version). We have added the following explanation in the text.

   Note that the towers with zero occupancy in the figure are known to have had issues and are masked, not showing up in the ML quality plot. The two aforementioned towers that are spotted from the ML quality plot exhibit some noticeable patterns in Fig.6(b).

5. **Your anomaly tagging threshold is based on fake anomalies. How is this reasonable? Does the distribution of fake anomalies match the distribution of real anomalies? Please mention.**

A: The fake anomalies are indeed a good representation of the real anomalies which can manifest in the detector as a single or group of zero-occupancy and/or hot towers.
So by being able to correctly tag these two anomalies at the smallest granularity of a single tower we do essentially cover all possible combinations of real anomalies that can be found in the occupancy plots from the detector.
We also derive a separate anomaly threshold for each anomaly scenario, so there's no assumption that anomaly of type A happens more often than type B.
And in the end, we choose a single best threshold for each sub detector that can catch all types of anomalies we test in the “fake anomaly” validation. Using this single best threshold we also test on real anomalies of various shapes and kinds and demonstrate that we can detect all of them, as described in Section 4.3 and Figure 5.