Machine Learning approach to CMS RPC HV scan data analysis

Mihaela Pehlivanova* on behalf of the CMS Muon group

University of Sofia

mihaela.pehlivanova@cern.ch*

1. Introduction

THE High Voltage (HV) scan is a critical calibration procedure performed at the start of each data-taking year at the CERN LHC. This scan is essential for ensuring the proper functioning of RPC detectors in the CMS experiment by establishing their correct working points. Because some part of the analysis is performed manually, an ML-based tool has been developed to speed up the process of analysing RPC HV scan data. This tool utilizes an artificial neural network (ANN) trained to discard outliers and accurately approximate efficiency behavior, even when data is missing within the range where efficiency typically stabilizes (efficiency plateau).

2. RPC HV scan data analysis

Efficiency is measured for every HV point and the resulting distributions are fitted using a sigmoid function defined as:

$$\epsilon = \frac{\epsilon_{max}}{1 + e^{-slope_{50\%}(HV - HV_{50\%})}} \quad (1)$$

where $slope_{50\%}$ characterize the slope of the sigmoid, ϵ_{max} represents the plateau of the distribution, and $HV_{50\%}$ is the value of the voltage at 50% of the maximum efficiency [1]. The working points per roll (WP_{roll}) for the RPCs in the barrel and endcap regions are defined as $HV_{95\%} + 100$ V and $HV_{95\%} + 120$ V, respectively. A total 480 and 293 HV channels in barrel and endcap regions, respectively, supply between 2 barrel and 6 endcap rolls. The working point per channel is defined by the following [1]:

$$WP_{CH} = \begin{cases} \langle WP_{roll} \rangle & \text{if } WP_{roll}^{\text{Max}} - WP_{roll}^{\text{Min}} \leq 100 \text{ V} \\ WP_{roll}^{\text{Min}} & \text{if } WP_{roll}^{\text{Max}} - WP_{roll}^{\text{Min}} > 100 \text{ V} \end{cases} \quad (2)$$

3. Fourier Space Autoencoder (FSAC)

A dedicated autoencoder ANN has been developed to operate in Fourier space, approximating efficiency versus HV curves for each double gap.

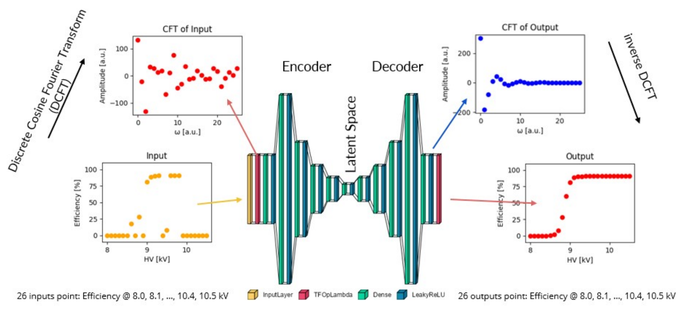

Figure 1: *Diagram of Fourier Space Autoencoder. The model consists of an input layer, a DCFT layer, an autoencoder, an output layer in the reciprocal space domain, and a subsequent layer that transforms the data back to the HV domain via inverse DCFT. The insets illustrate examples of inputs and outputs, showcasing the transformation between the HV and reciprocal space domains.*

4. Training and validation

A synthetic dataset, derived from a sigmoidal function, was used for training and validation purposes, with 80% of the data designated for training and 20% for validation. The machine learning model was trained and validated, yielding a 3% error rate during testing.

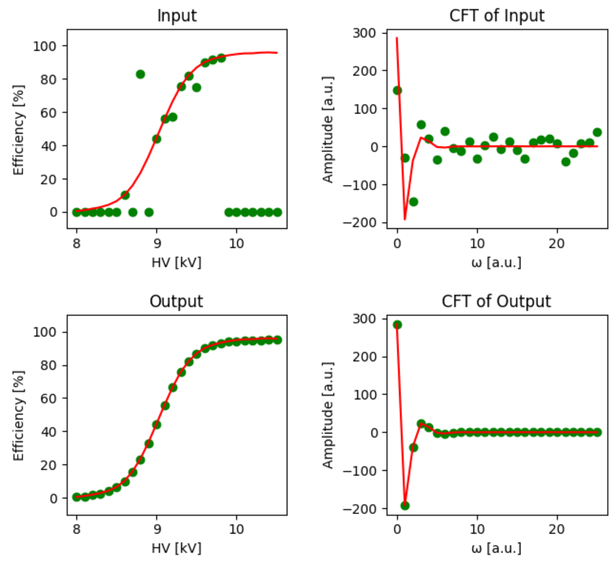

Figure 2: *Examples of model training input with outliers and the corresponding output. In the HV domain plots, the red line represents the original function from which the data was generated. In the reciprocal space domain plots, the red line shows the Cosine Fourier Transform (CFT) of the function, illustrating how the original function is broken down into its real (cosine) frequency components. On the top row the green dots correspond to the generated data points and their respective CFT. The bottom row showcases the model's output in both the HV and reciprocal space domains.*

5. Performance on data

The model has been tested on RPC HV scan data from 2024, where it uses measured efficiency values across different supply voltages to estimate detector efficiency within the range of 8.0 to 10.5 kV. It generates 26 hv-efficiency points, which we fit by sigmoidial function (eq.(1)) to obtain the fitting parameters.

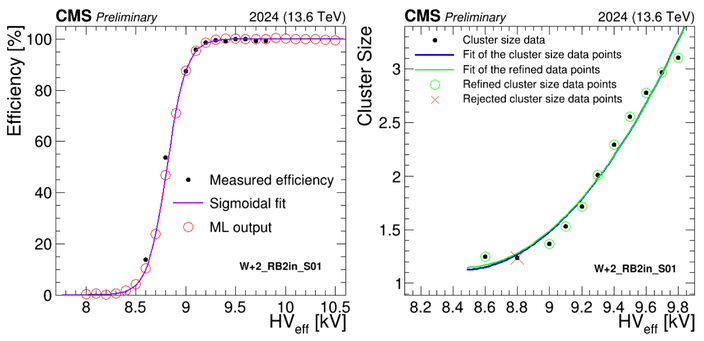

Figure 3: *Efficiency and cluster size curves after applying FSAC for case of good efficiency curve. The chamber is W+2 RB2in S01 Backward. The data point at 8.8 kV is excluded from the cluster size plot because the absolute difference between the predicted and actual efficiency values at that voltage exceeded the associated error.*

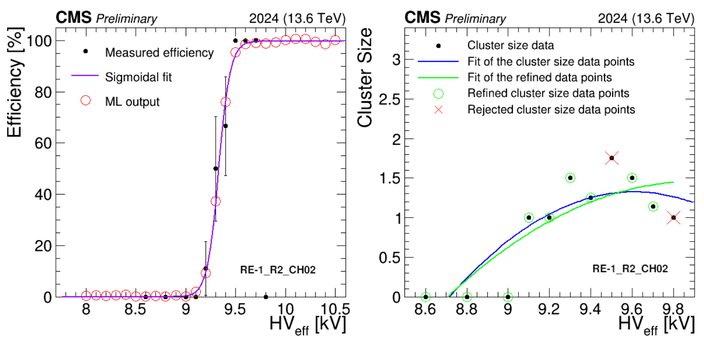

Figure 4: *Efficiency and cluster size curves after applying FSAC for case where the last point has zero efficiency. The chamber is RE-1 R2 CH02 A. The data points at 9.5 kV and 9.8 kV were excluded from the cluster size plot because the absolute difference between the predicted and actual efficiency values at these voltages exceeded the associated errors.*

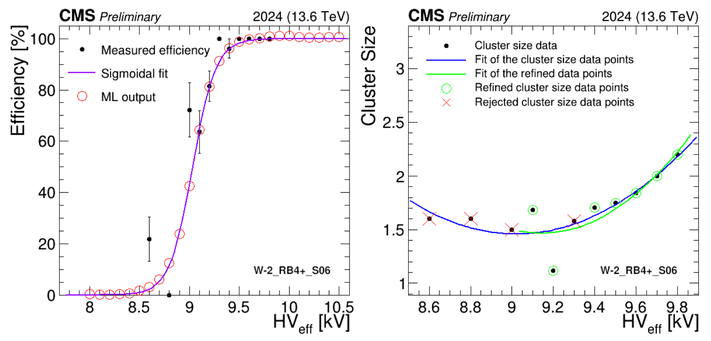

Figure 5: *Efficiency and cluster size curves after applying FSAC for case where there are outliers present in the efficiency plot. The chamber is W-2 RB4+ S06 Backward. The data points at 8.6 kV, 8.8 kV, 9 kV and 9.3 kV were excluded from the cluster size plot because the absolute difference between the predicted and actual efficiency values at these voltages exceeded the associated errors.*

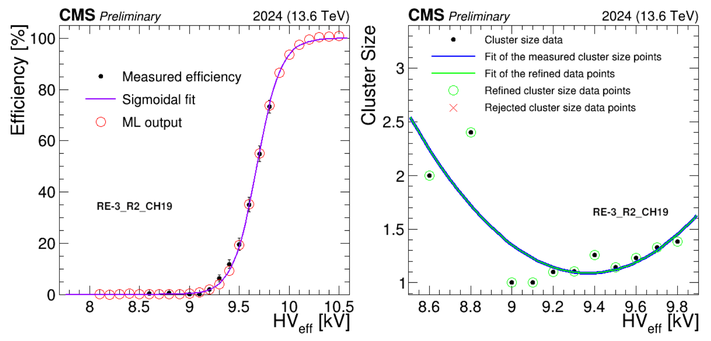

Figure 6: *Efficiency and cluster size curves after applying FSAC for case where the plateau is missing in efficiency curve. The chamber is RE-3 R2 CH19 A. .*

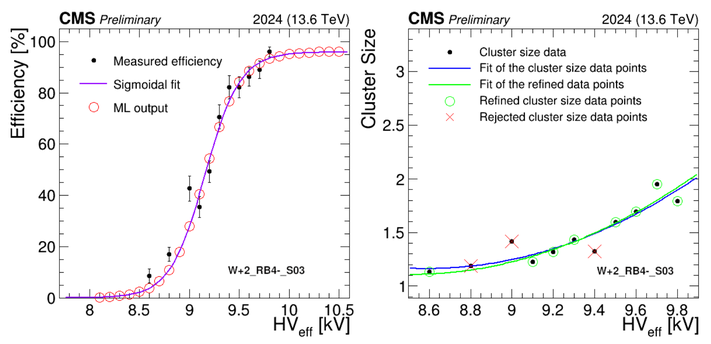

Figure 7: *Efficiency and cluster size curves after applying FSAC for case where there are outliers and missing plateau present in the efficiency plot. The chamber is W+2 RB4 S03 Backward. The data points at 8.8 kV, 9 kV and 9.4 kV were excluded from the cluster size plot because the absolute difference between the predicted and actual efficiency values at these voltages exceeded the associated errors.*

Data are taken with pp-collisions, and effiency and cluster size are evaluated using Segment Extrapolation Method. The model accurately estimates the parameters for defining new working points for each roll and chamber, effectively handling problematic data by automatically detecting and discarding outliers, eliminating the need for manual intervention and refitting.

6. Conclusion

The proposed machine learning algorithm can greatly speed up the analysis of calibration data from the CMS RPC HV scan. What used to take over three months can now be done in less than a week, making the process much more efficient.

References

[1] R. Reyes-Almanza *et al.*, *High voltage calibration method for the CMS RPC detector*, *Journal of Instrumentation* **14** (2019) C09046, 10.1088/1748-0221/14/09/C09046. Available at: https://dx.doi.org/10.1088/1748-0221/14/09/C09046.