Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

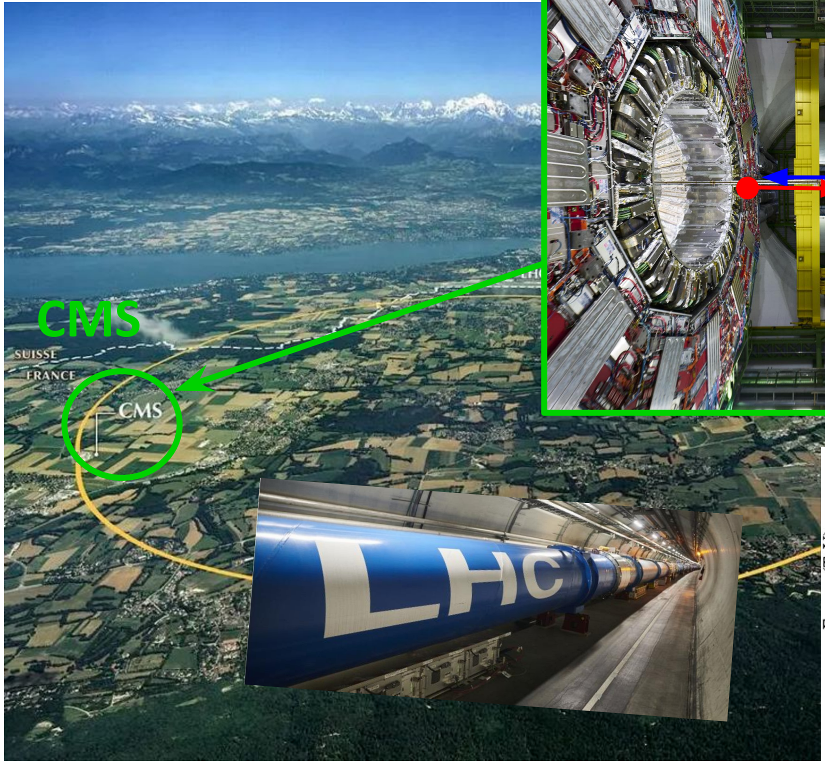
Anomaly detection for data quality monitoring of the Muon system at CMS

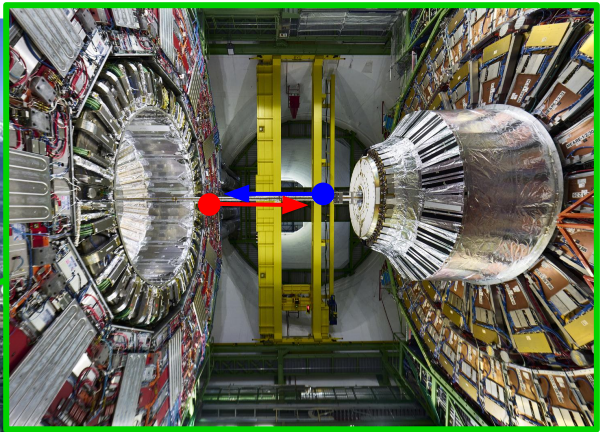
Marco Buonsante[1], Marco Cruciani[2], Federica Simone[3], Rosamaria Venditti[1]
on behalf of the CMS Collaboration

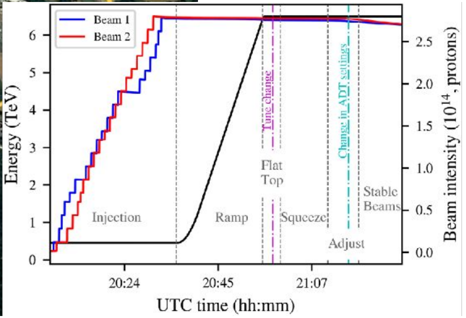
[1] Università e INFN Bari
[2] Università di Bologna
[3] Politecnico e INFN Bari

CHEP 2024, 19-25 October 2024, Krakow, Poland

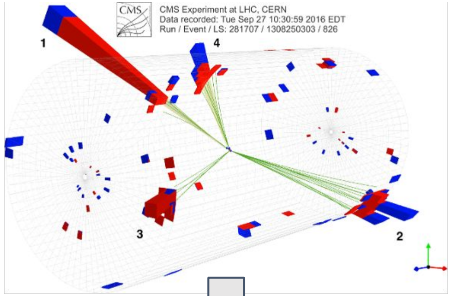
Data taking at Compact Muon Solenoid (CMS)







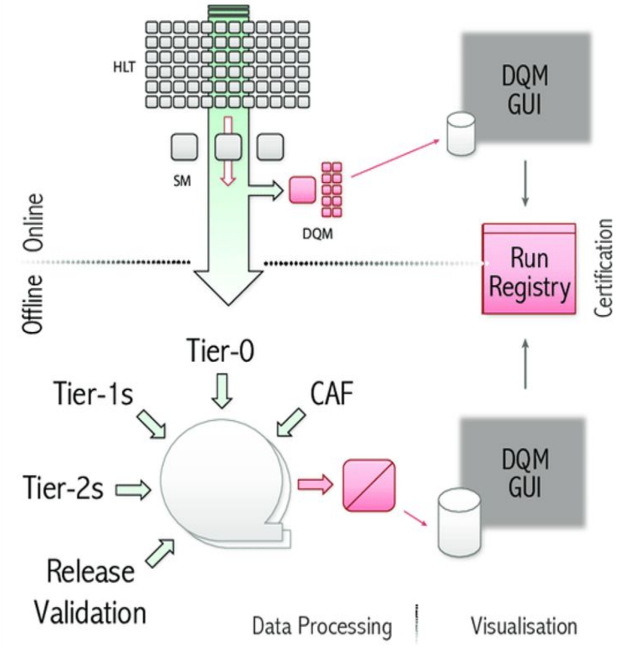
Recorded event



23 SEC **Lumisection (LS)**

"Run": thousands of LS

Data Quality Monitoring (DQM)



Online DQM: Promptly raise alarms during data taking

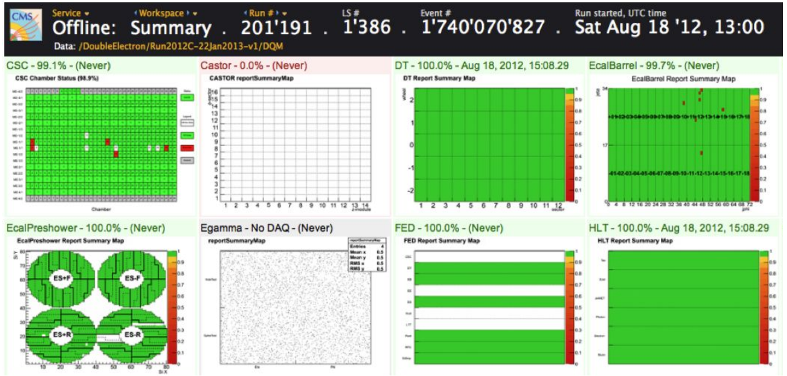
Offline DQM: Provide inputs to experts for investigating issues

Offline Data Certification (DC): Certify high-quality data for physics analysis

Online ME: Histograms are filled in real time during the run.

Offline ME: Histograms are available for inspection with per-run granularity

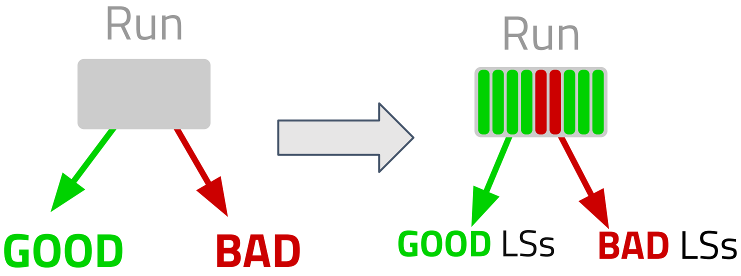
Monitoring Elements (ME):



[J. Phys.: Conf. Ser. 513 032024]

DQM challenges and limitations

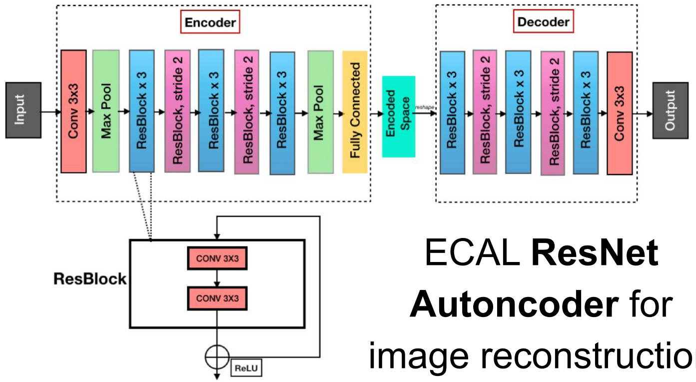
- Online monitoring is a **highly time-sensitive** operational task
- Data Certification should **ensure high quality data** while **limiting false positive rate**



- Limited **time granularity** (run) may hide transient issues affecting only a few lumisections
- Drawback: **per-LS approach** increases the number of MEs by a factor $\mathcal{O}(10^3)$
  - Human inspection not feasible

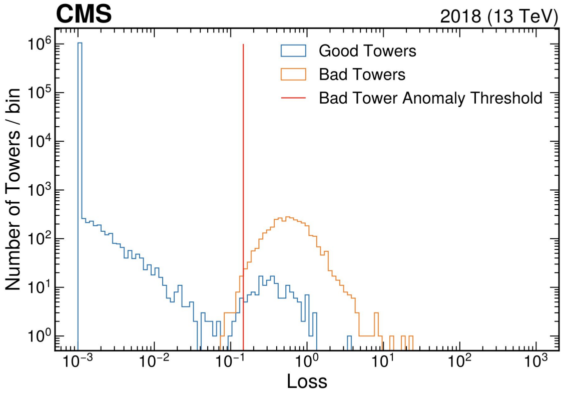
State of the art

- Ongoing effort within CMS in using ML to **automatise the monitoring** of different subsystems:
  - **ECAL** [Comput Softw Big Sci 8, 11 (2024)]: Online DQM,
  - Tracker [CERN-CMS-DP-2024-070]: DC,
  - HCAL [Sensors 2023, 23(24), 9679] Online & Offline DQM

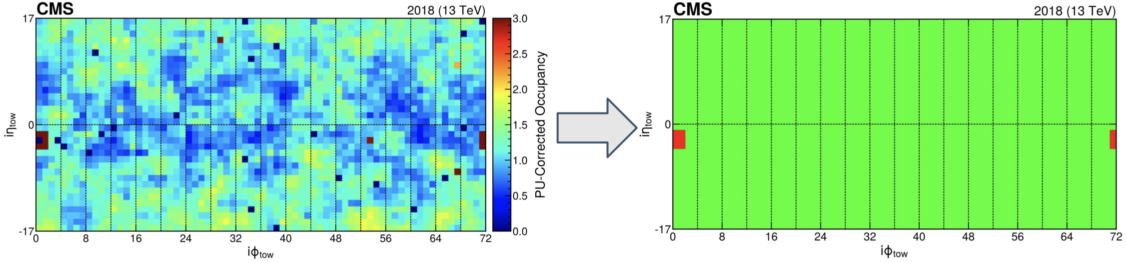


ECAL **ResNet Autoncoder** for image reconstruction

ECAL Loss Distribution



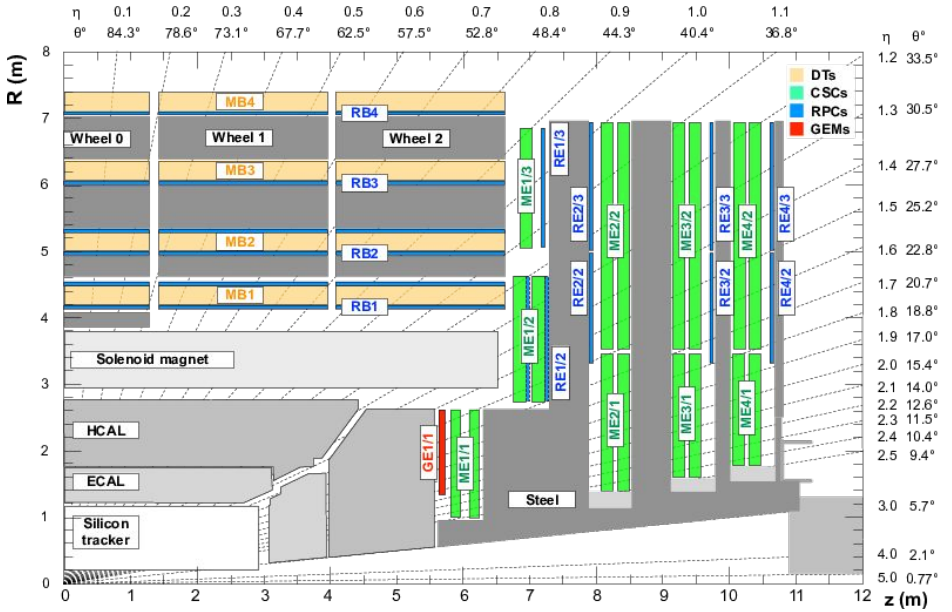
ECAL Barrel Occupancy



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

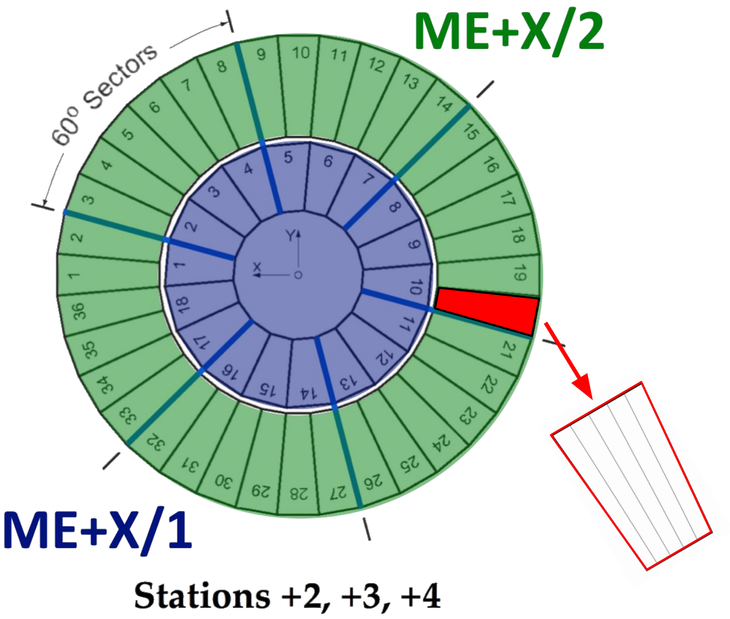
Anomaly detection for the online monitoring of the CMS Muon System

- **Goal**: Develop a flexible ML anomaly detection tool
- **Inputs**: Occupancy plots from muon system sub-detectors (taking inspiration from ECAL [Comput Softw Big Sci 8, 11 (2024)])
- **Challenges**: Low rate of particles reaching the muon system implies in low occupancy
- **Tested with Cathode Strip Chambers (CSC)** occupancy plots
  - **150k** LSs all certified as **GOOD** (from 2024 data taking)



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

CSC detector and expected anomalies



- CSCs are placed in the **endcaps** of the CMS muon system
- They are arranged in disks called **stations**
- Stations (from 2 to 4) have two **rings**: MEX/1 MEX/2
- MEX/1 (MEX/2) is divided in **chambers** of 20° (10°), each readout by **5 frontend boards**

- Possible anomalies manifest as a **deficit** or **excess** of entries in a chamber or a readout sector.

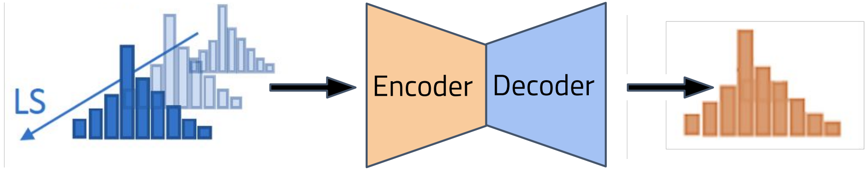
Workflow

1 – Data preprocessing

- Sum LSs ME distributions
- Taking in account data taking conditions (instantaneous luminosity, LHC status)

2 – Training

- ResNet Autoencoder for Image Reconstruction
- General idea: the model should learn an abstract representation of good data

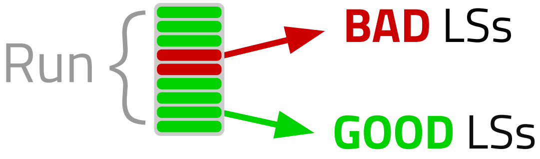


3 – Validation

- Measure performance on labelled data
- Set metrics and thresholds

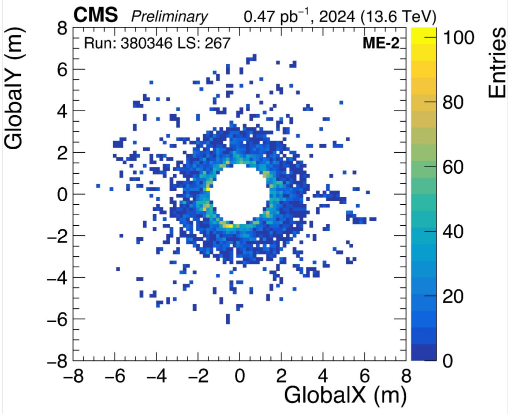
4 – Flag BAD/GOOD

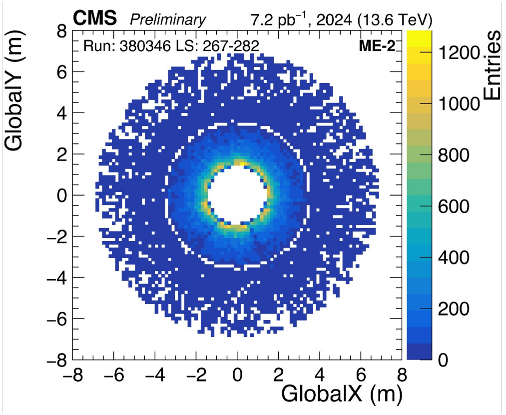
- Flag data with few LS granularity. Either reject anomalous data or further investigate

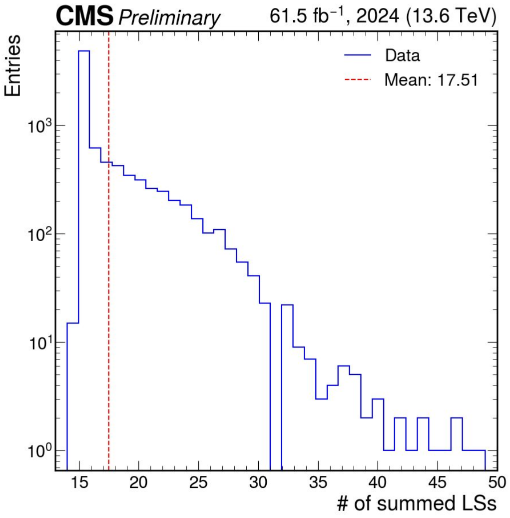


CSC data preprocessing (1)

- Data are filtered requiring **GOOD** LSs with ist. lumi > 2 nb^{-1} s^{-1}
- Muon System is characterised by **low occupancies** (especially for MEX/2) → **Can't** use a single LS
  - Sum over consecutive LSs up to int. lumi of 7 pb^{-1}





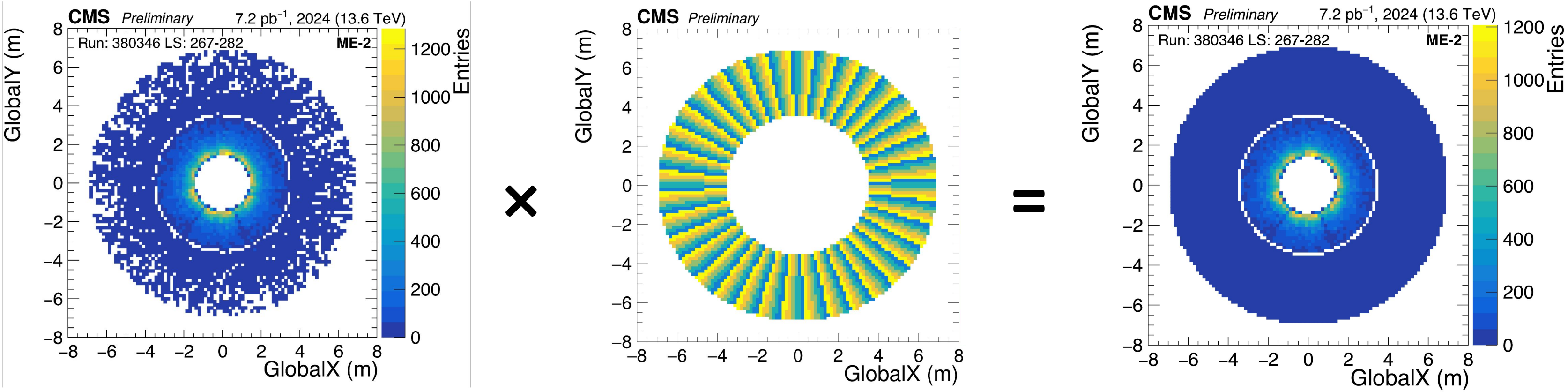


Distribution of the number of summed LSs

CSC data preprocessing (2)

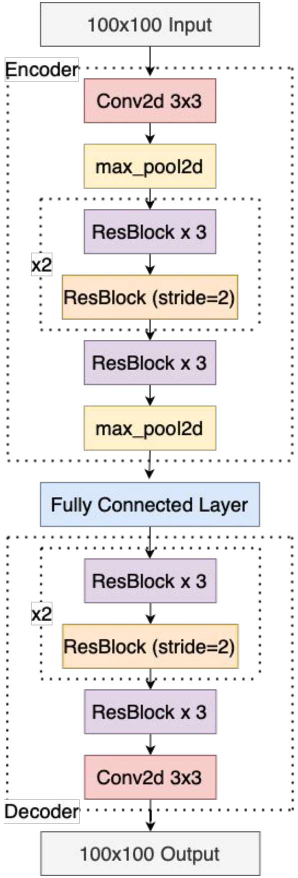
After merging consecutive LS, the images still have **empty regions** at low η (MEX/2).

- The statistical fluctuations can affect the generalization capability of the autoencoder.
- The external part of each image is **rebinned in slices** along the ϕ coordinate (following the geometry of the detector readout)



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Training



ReCaS BARI

[ReCaS](#)

NVIDIA

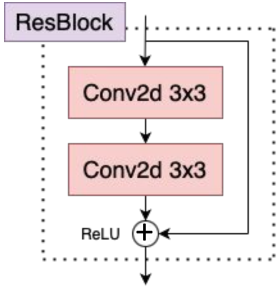
PyTorch

Model: ResNet for unsupervised reconstruction of images

Input: ~8000, 100x100, preprocessed CSC occupancy images

Training loss: [L1Loss](#)

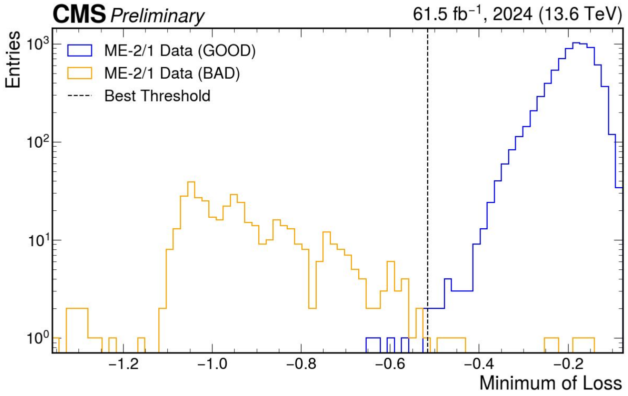
Strategy: Two independent trainings for MEX/1 and MEX/2



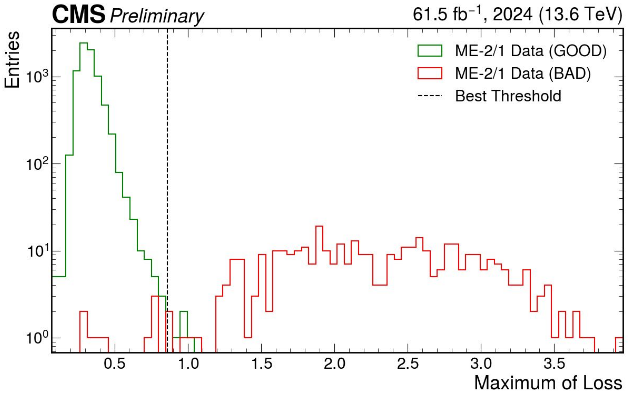
Validation: MEX/1

- A set of good and problematic images is isolated from data by visual inspection.
- Compute the “Loss” as signed difference between input- and reco- image divided by mean input images
- Study the distribution of the minimum and maximum of the Loss
- Threshold → optimizing **F1 score**

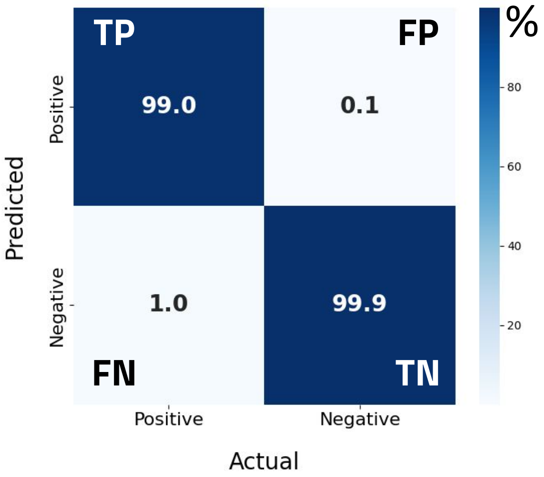
Minimum



Maximum



Confusion Matrix

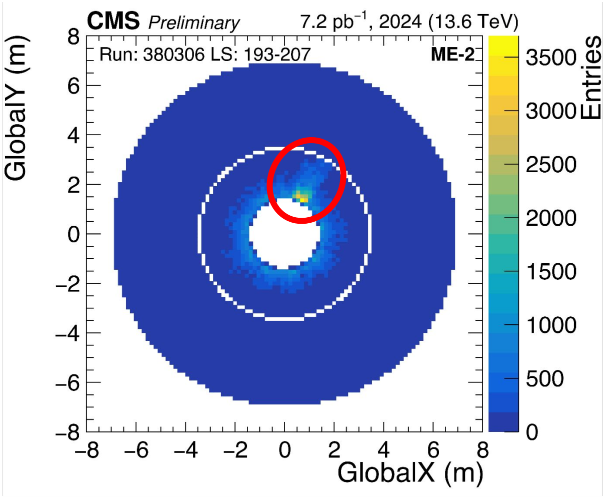


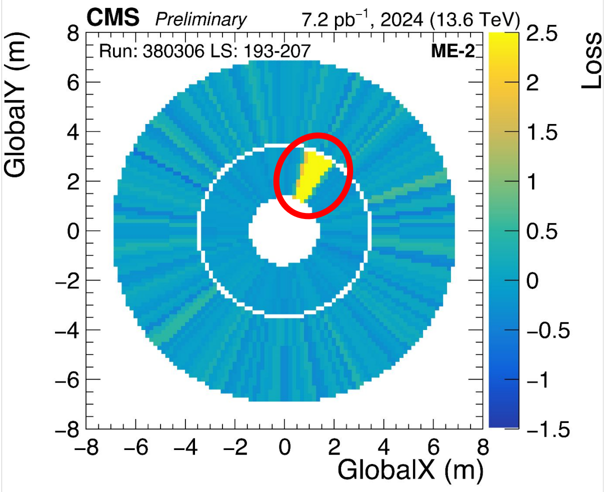
Positive = Anomaly

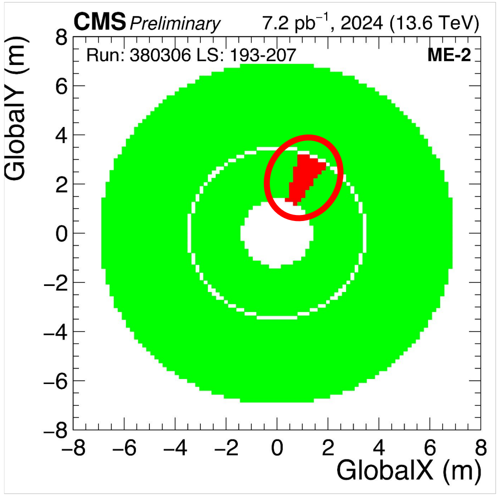
Flagging: MEX/1

- Anomaly found in data, clear **hotspot** visible data
- The algorithm provides a map showing the anomalous region

```
May 02, 2024
Run 380306
LS: 193 to 207
```



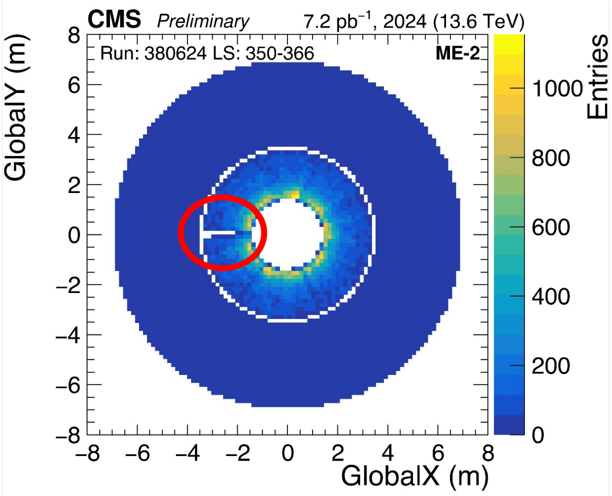


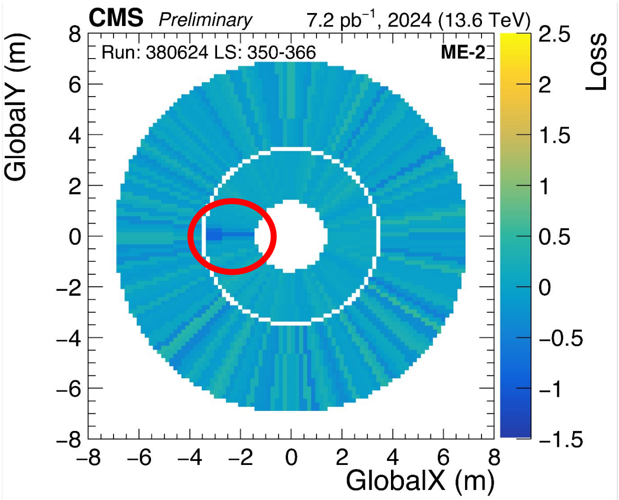


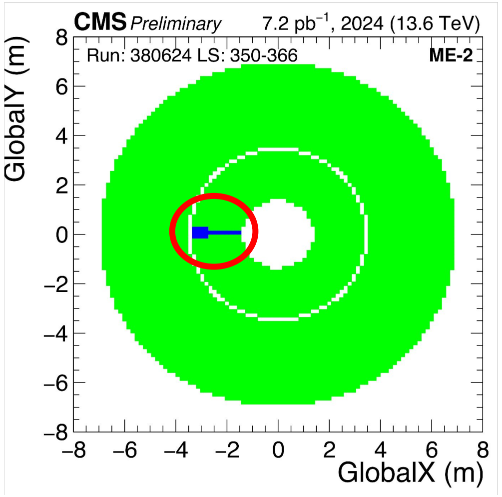
Flagging: MEX/1

- Anomaly found in data, clear **empty region** visible in data
- The algorithm provides a map showing the anomalous region

May 11, 2024
Run 380624
LS: 350 to 366





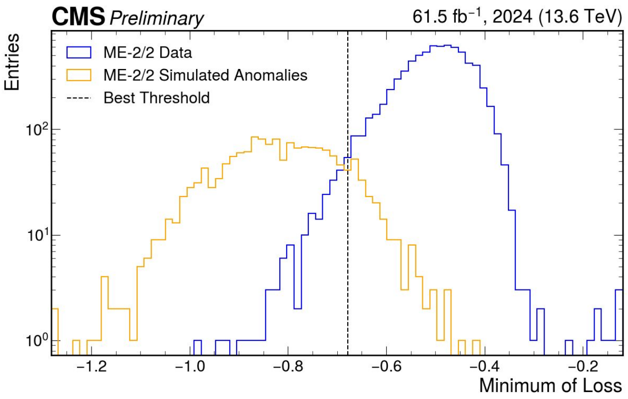


Validation: MEX/2

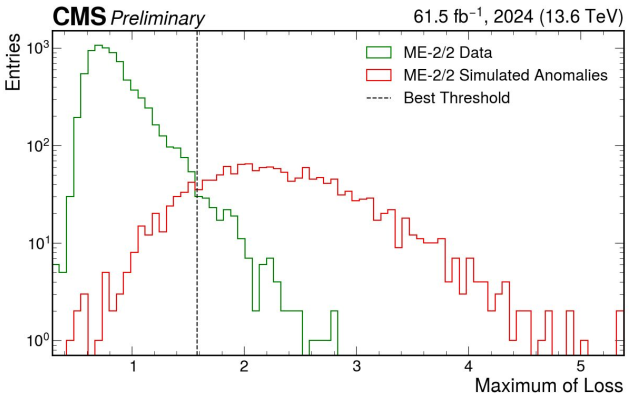
- Low statistics → Difficult to isolate anomalies by eye
- Generate artificially anomalies introducing over/under fluctuations in slices of the detector (based readout sectors geometry)

- Compute the "Loss" as difference between input- and reco- image divided by mean input images
- Threshold → optimizing **F1 score**

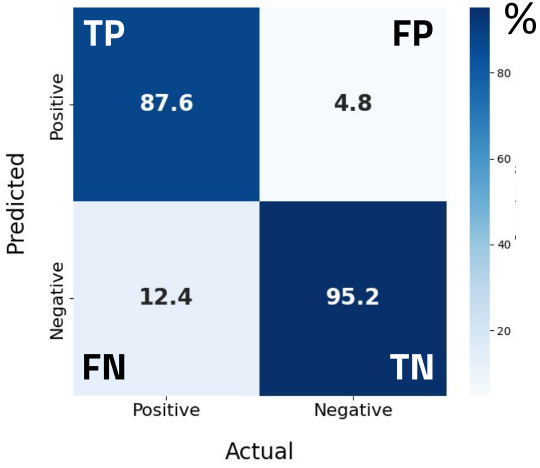
Minimum



Maximum



Confusion Matrix

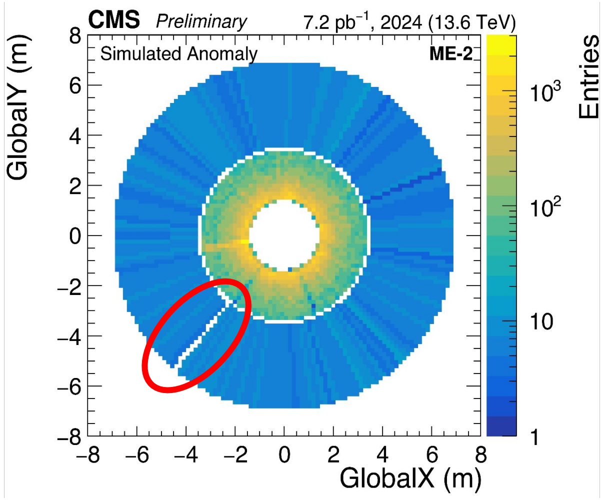


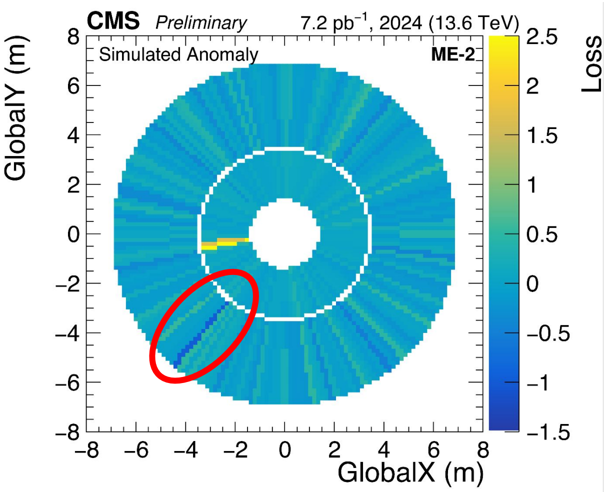
Positive = Anomaly

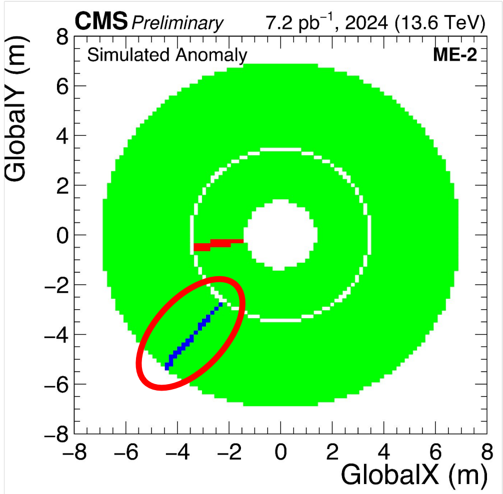
Flagging: MEX/2

- Simulated anomaly, **empty region** visible in input image + additional **hotspot** in ME-2/1
- The algorithm provides a map showing the anomalous region

Simulated Anomaly



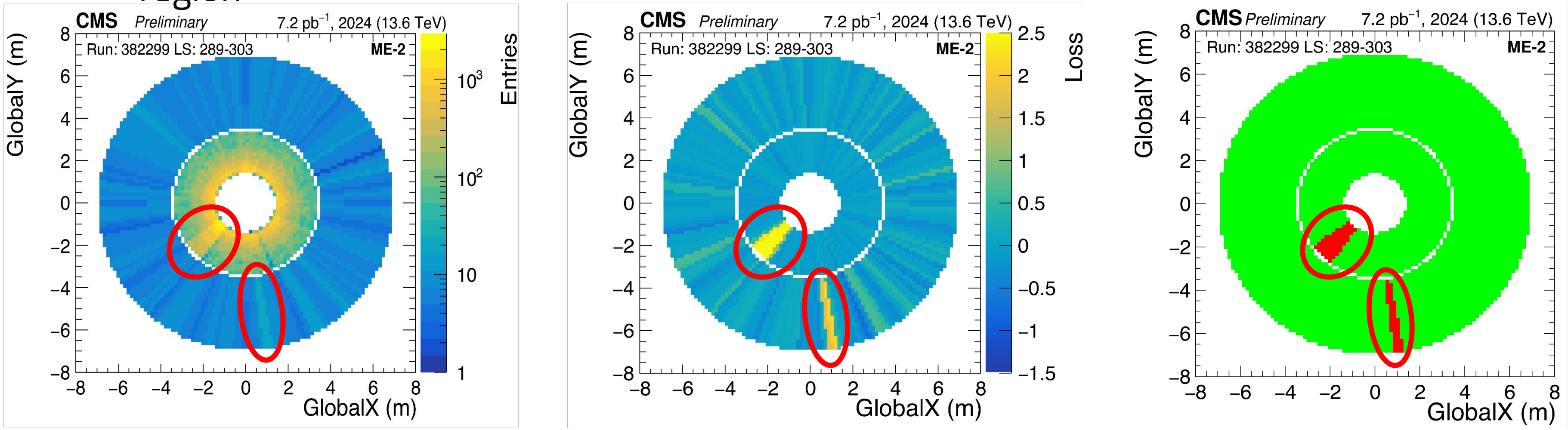




Flagging: MEX/2

- Anomaly found in data, **hotspot** visible in data for both ME-2/1 and ME-2/2
- The algorithm provides a map showing the anomalous region

June 21, 2024
Run 382299
LS: 289 to 303



Conclusions

- To showcase the applicability of **ML-based** monitoring to the CMS Muon System an **anomaly detection tool** has been developed for the **online monitoring of the CSC** detector with time granularity of the order of **few lumisections**.
- Anomalies are identified in the **2D occupancy maps** of reconstructed hits in the CSC ME2-4 stations.
- The algorithm is trained on a set of images from certified data, each corresponding to about 15 LS, separately for ME-2/1 and ME-2/2 rings.
- After training, a strategy is defined for **identifying anomalous images**. The fraction of anomalous images correctly labeled by the algorithm is **99% for** ME-2/1 and above **85%** for ME-2/2 (optimization still in progress).

This work is partially supported by ICSC – Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing, funded by European Union – NextGenerationEU

Thanks for your attention!