Automating ATLAS control room anomaly detection with machine learning

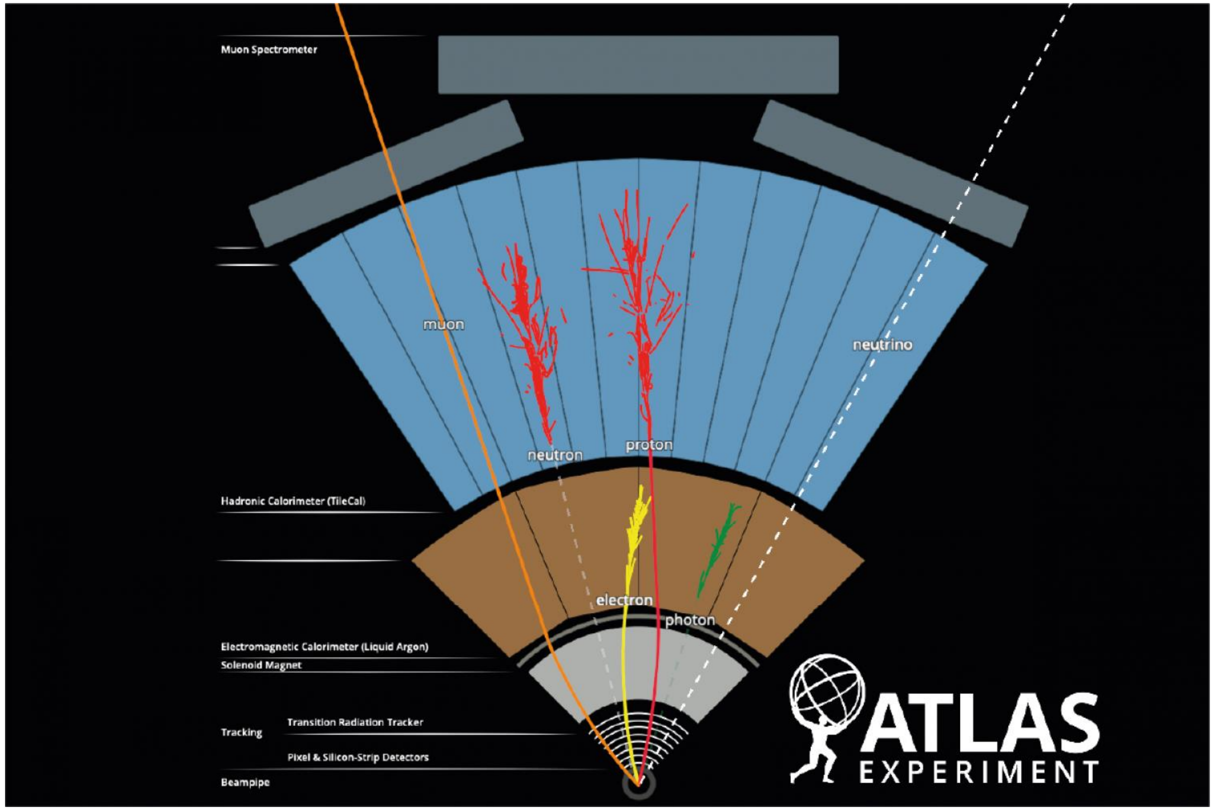
Avery Hanna

Supervisors: Mario Campanelli, Antoine Marzin

August 7, 2024

ATLAS data acquisition

- Detector composed of many subdetectors that allow us to identify and reconstruct the paths of different particles
- L1 trigger — first step of choosing what data to keep
- Data from these different subdetectors read out to control room where people review for anomalies



Motivations

- High-quality data acquisition via effective anomaly detection in ATLAS operation
- Currently, shifters manually monitor data in control room
  - Problems in consistency, accuracy
  - Shifters monitor many plots over 8-hour period
  - High personnel demand

Approach:

Online machine learning model to watch incoming time-series data and alert staff to anomalies

Data

Features (normalized by pileup):

- L1 trigger rates: frequency with which detect electrons, muons, missing transverse energy, jets
    - Multiple L1 rates for each corresponding to different energy levels
- L1 muon sector logic inputs: muon rate by section of detector
- Pileup: average number of interactions per bunch crossing. Constant for ~1/2 run, then decays

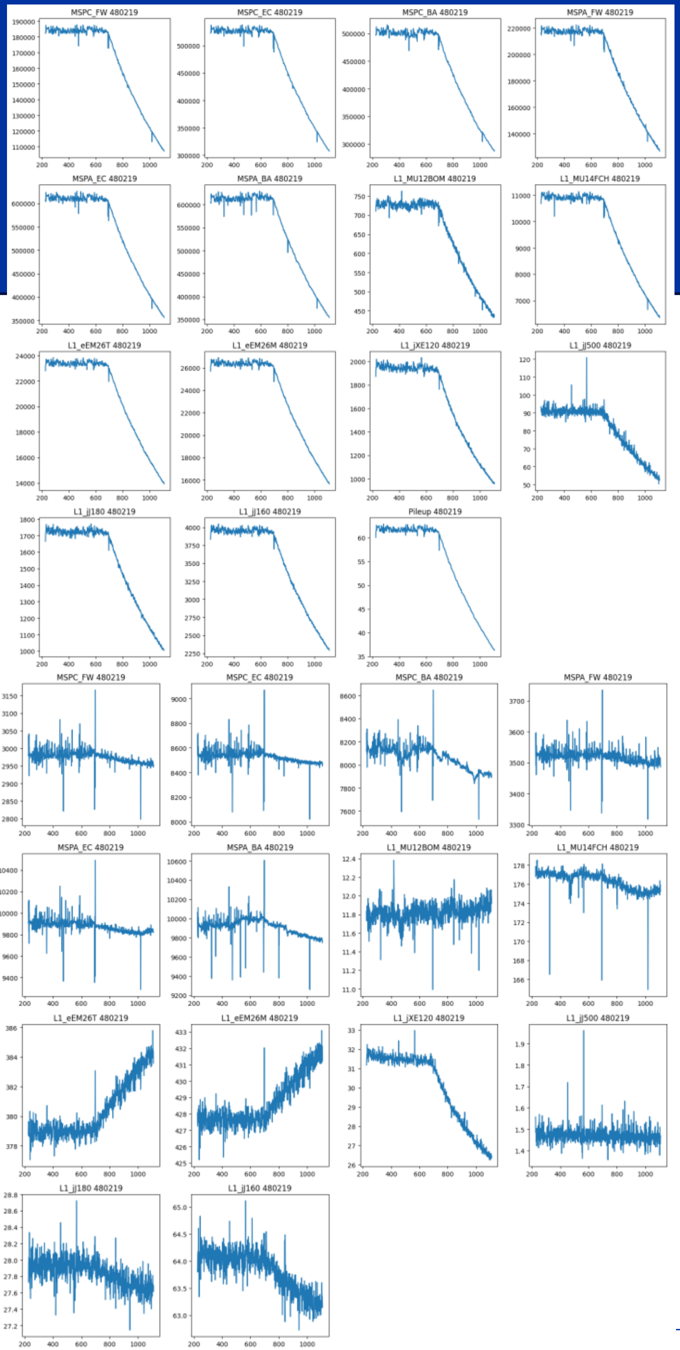
Input

- Time series data of 5 samples
- Each sample includes values for each of 14 features
- Robust scaling

Output

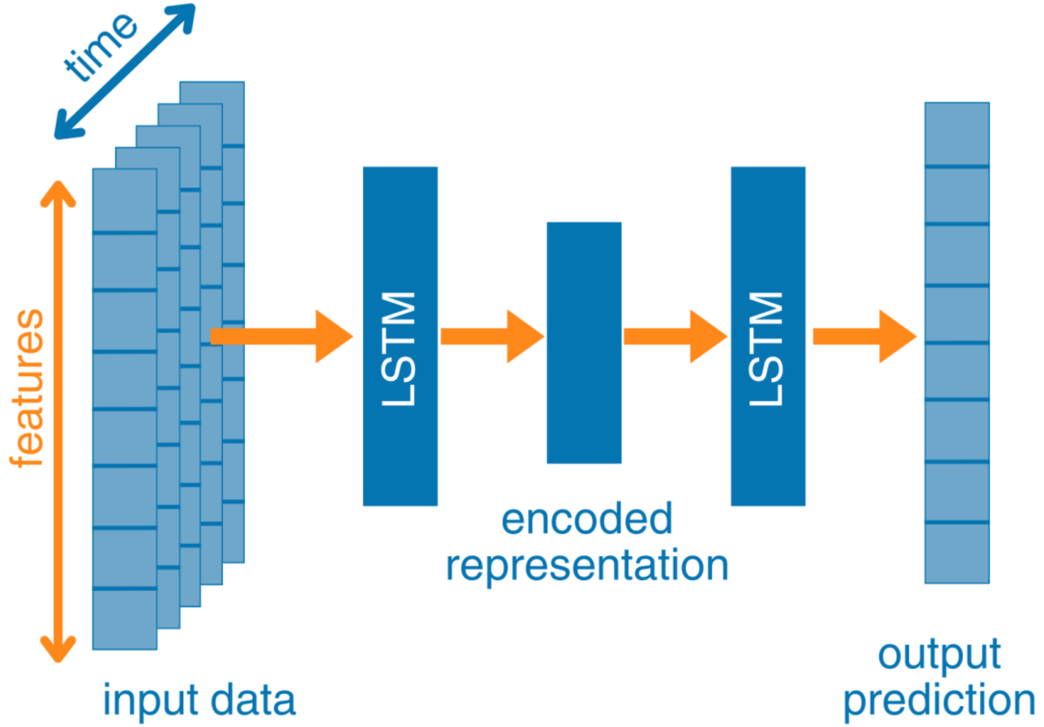
- Prediction of 14 features for next sample (one luminosity block/one minute in future)

Normalizing by pileup



Model architecture

- Model trained on clean data so low mean squared error (MSE) with clean values, high MSE with anomalies
- Set threshold MSE between clean and anomalous classifications
- Long short-term memory (LSTM) layers incorporate time-series element of data
- Autoencoder shape forces model to learn lower dimensional representation of data
    - Will be especially important as step up number of features
- Small network with only four layers



Model performance – loss, MSE

Training data: 10 full runs (around 8000 datapoints)

Test data: single full run (around 900 datapoints)

Huber loss

- Quadratic for small errors, linear for larger
- Chosen because don't want outliers in data to influence training as occurs with MSE

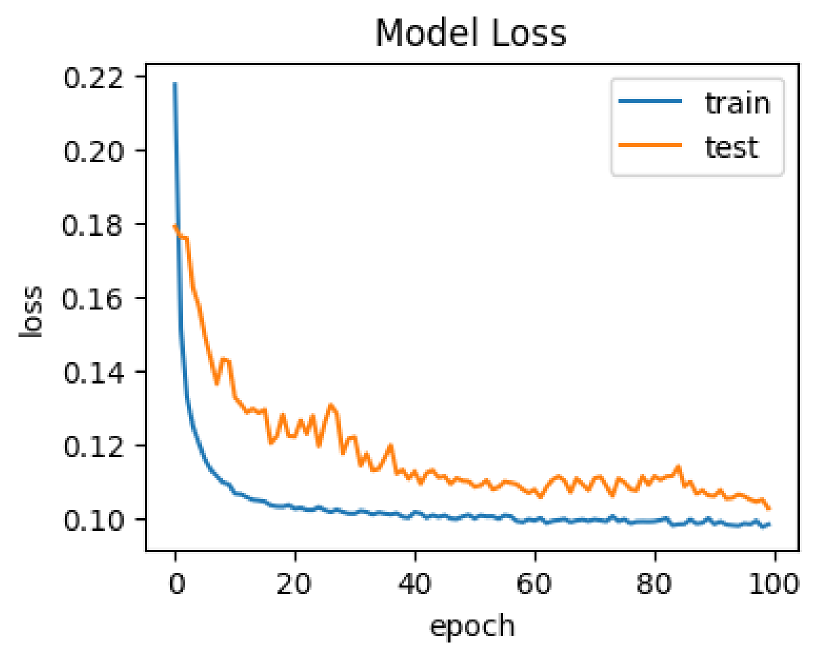
$$a = y - f(x)$$

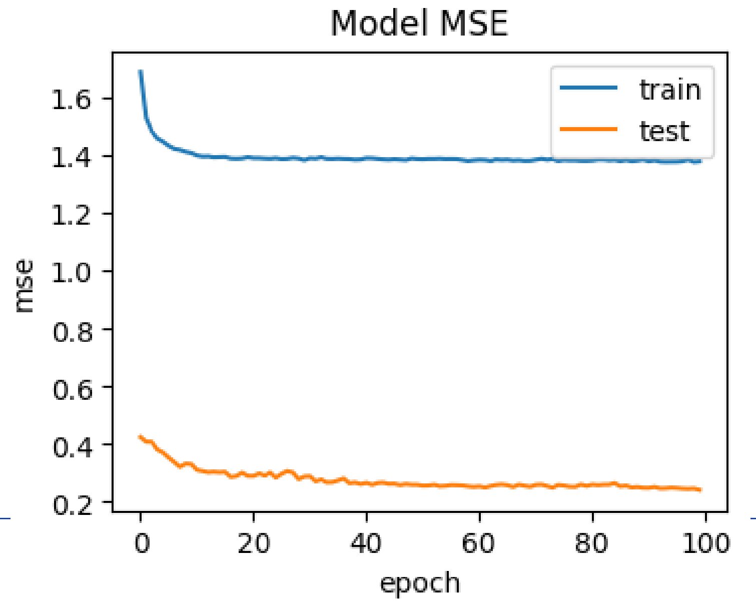
$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \le \delta, \\ \delta \cdot \left(|a| - \frac{1}{2}\delta\right), & \text{otherwise.} \end{cases}$$

Mean squared error

- A few outliers cause large separation in train and test

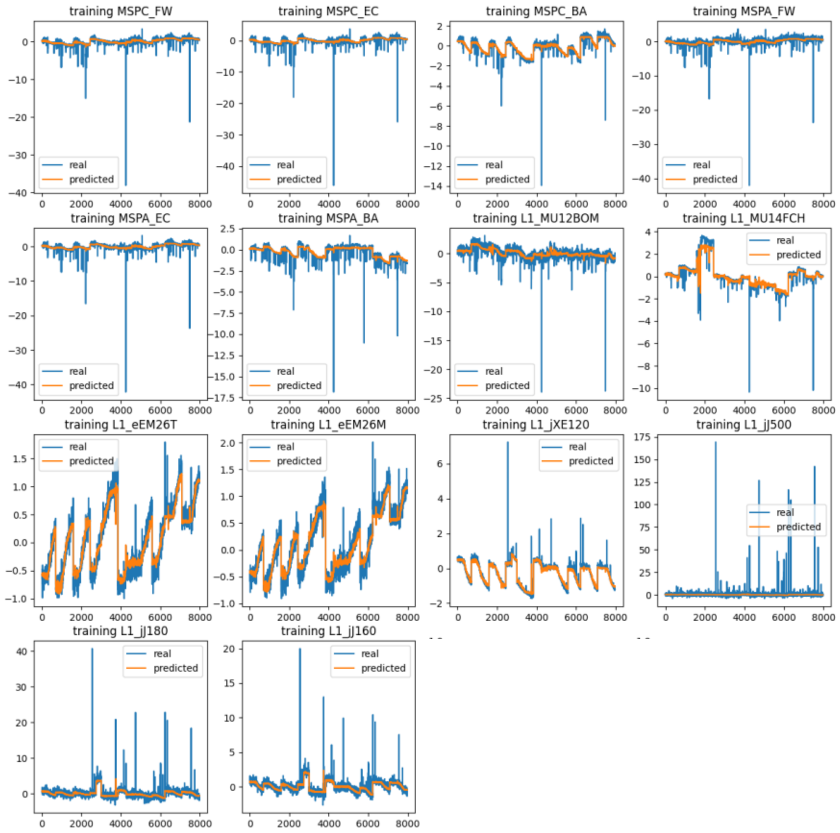
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$



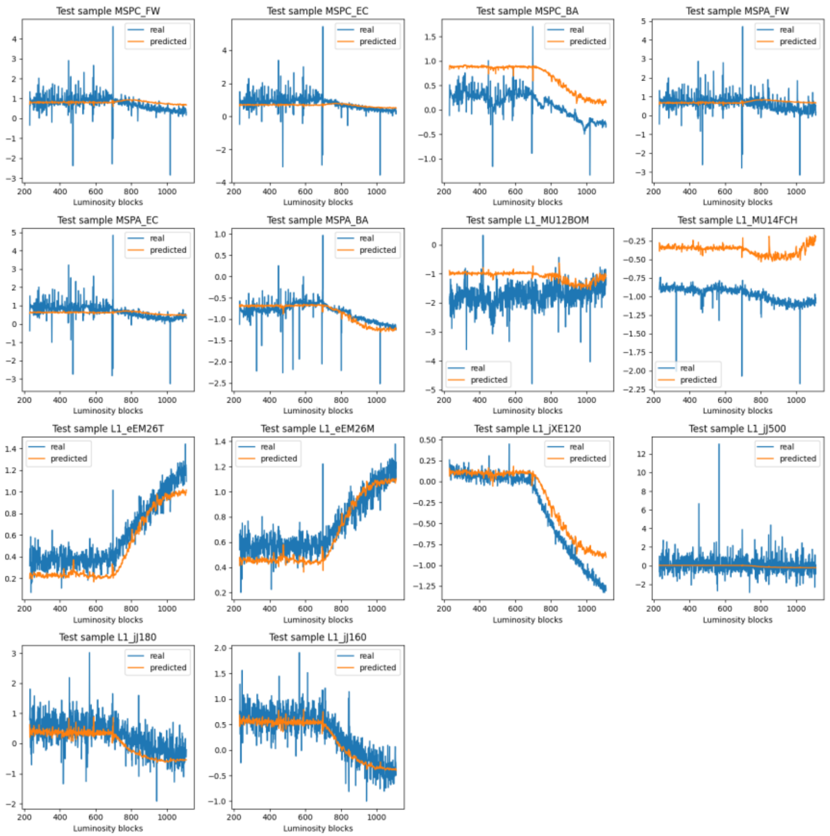


Model performance – training and test predictions

Predictions align well with training data, smooth out variations

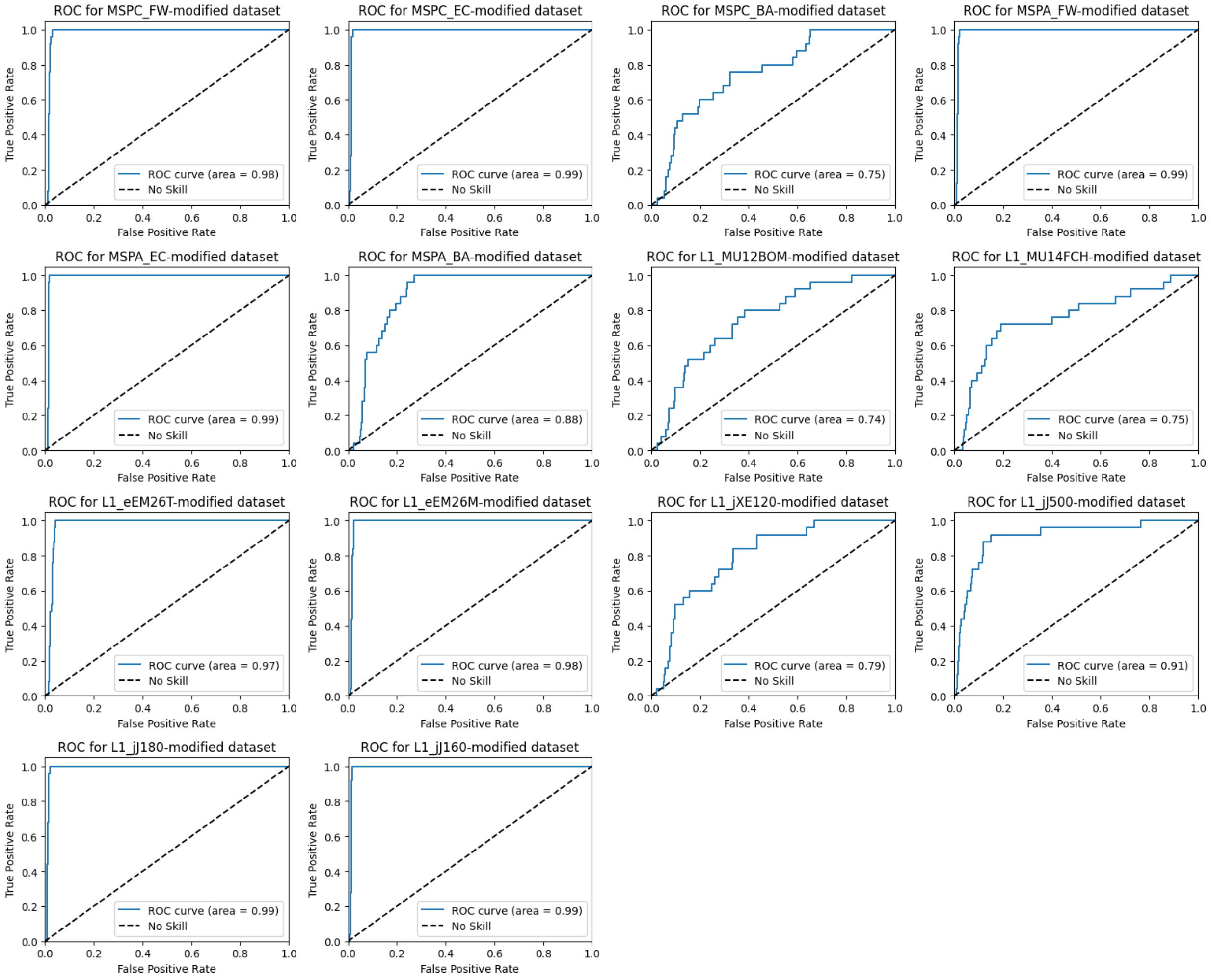


Observe offset between prediction and real in certain features, worse performance in tail of run



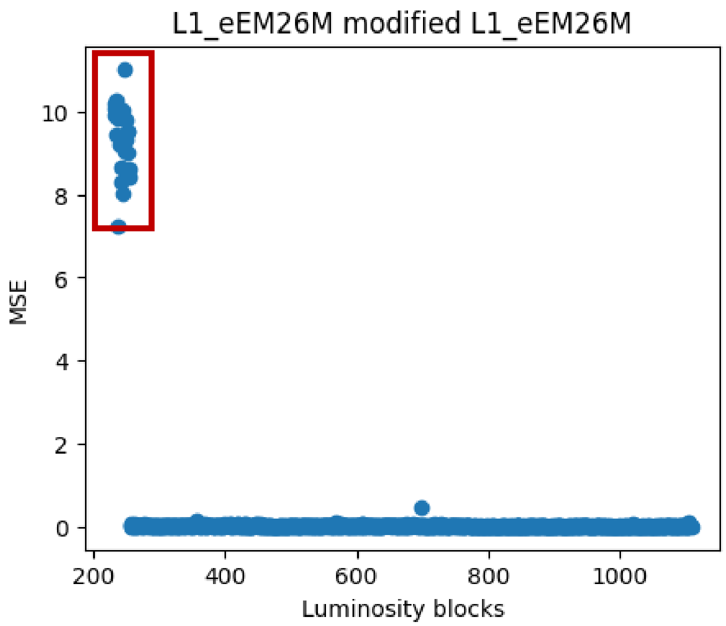
Model performance – artificial 5% anomaly

- For single feature in test data set, first 30 LBs increased by 5%
- On right, ROC curves show how performance differs based on which feature was modified
- AUC varies from 0.74 for L1_MU12BOM-modified dataset to 0.99 for five datasets

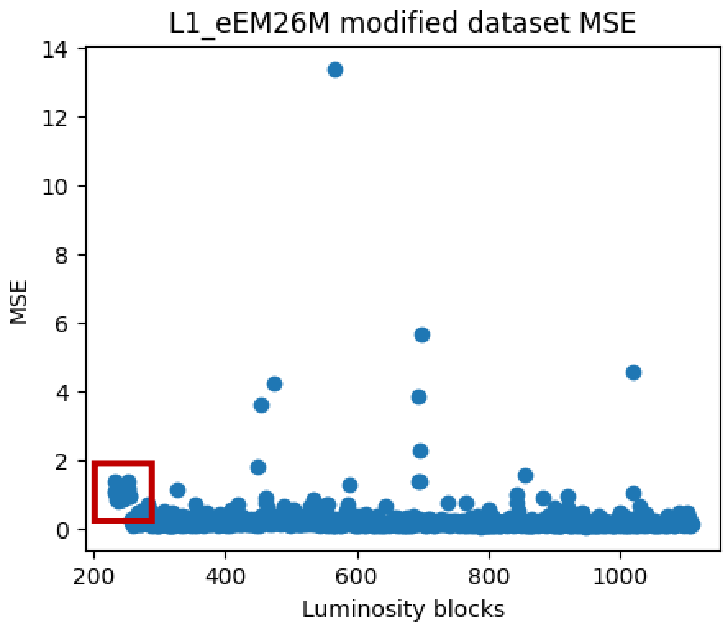


Model performance – artificial 5% anomaly

Looking at MSE for the specific feature that was modified, see clear spike



After averaging over all features still see bump, but not as clearly distinguished

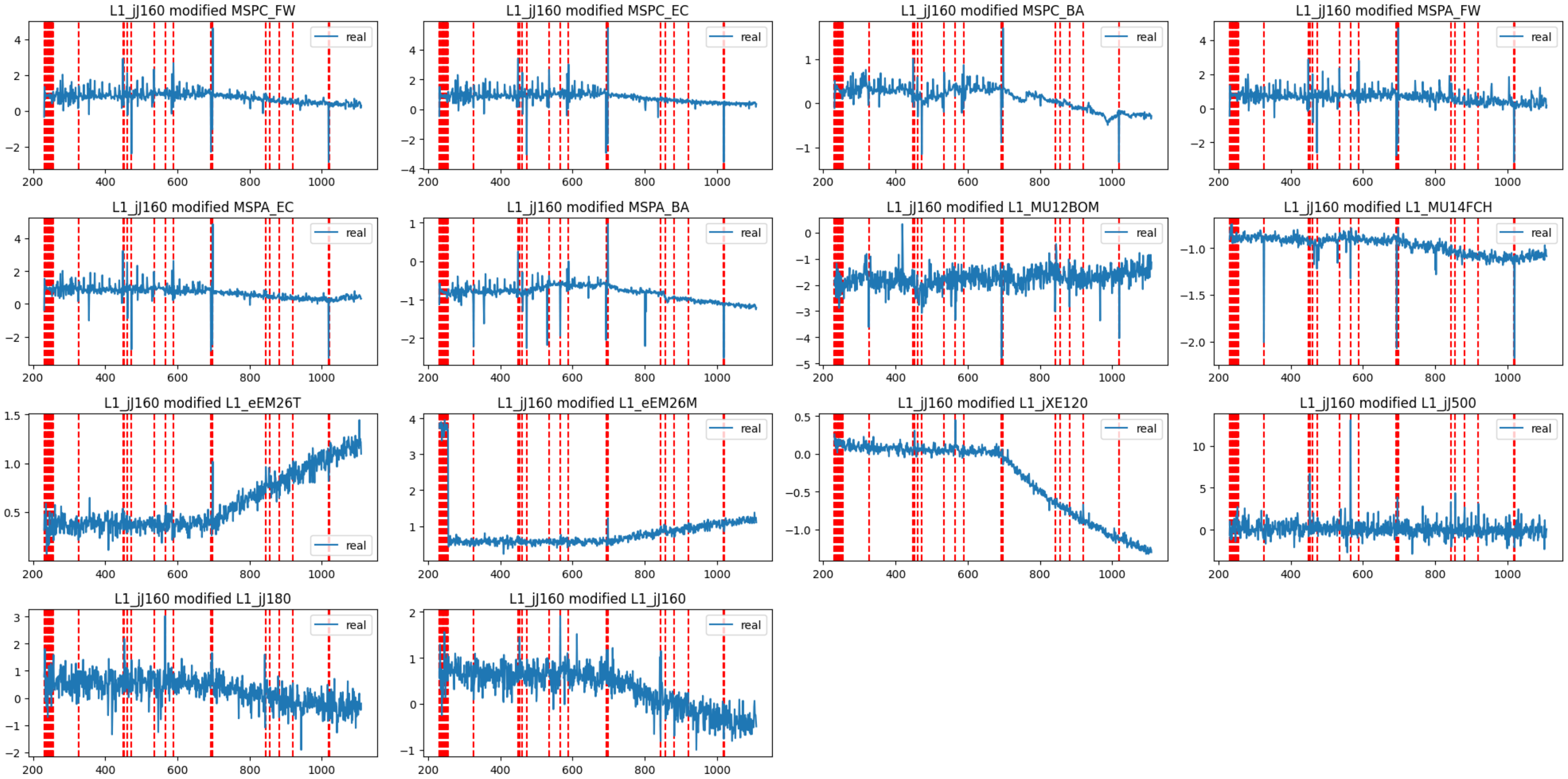


Model performance – artificial 5% anomaly

24/25 = 96% modified LBs correctly detected as anomalies

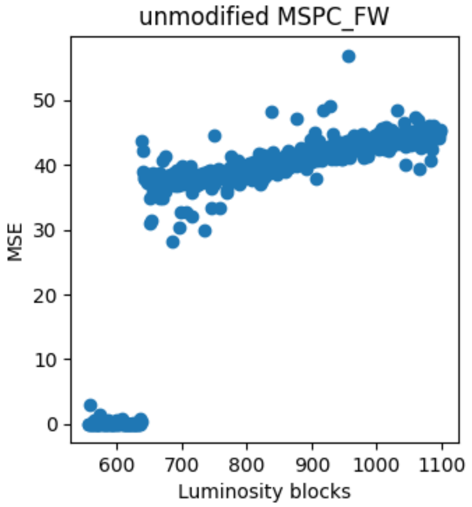
19/859 = 2.2% clean LBs falsely classified as anomalies

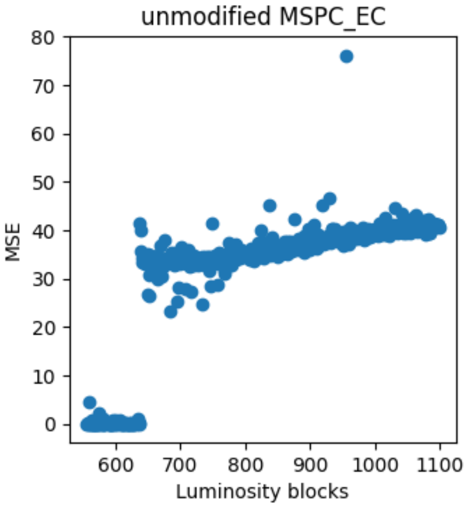
- Several of these correspond to real spikes in the dataset

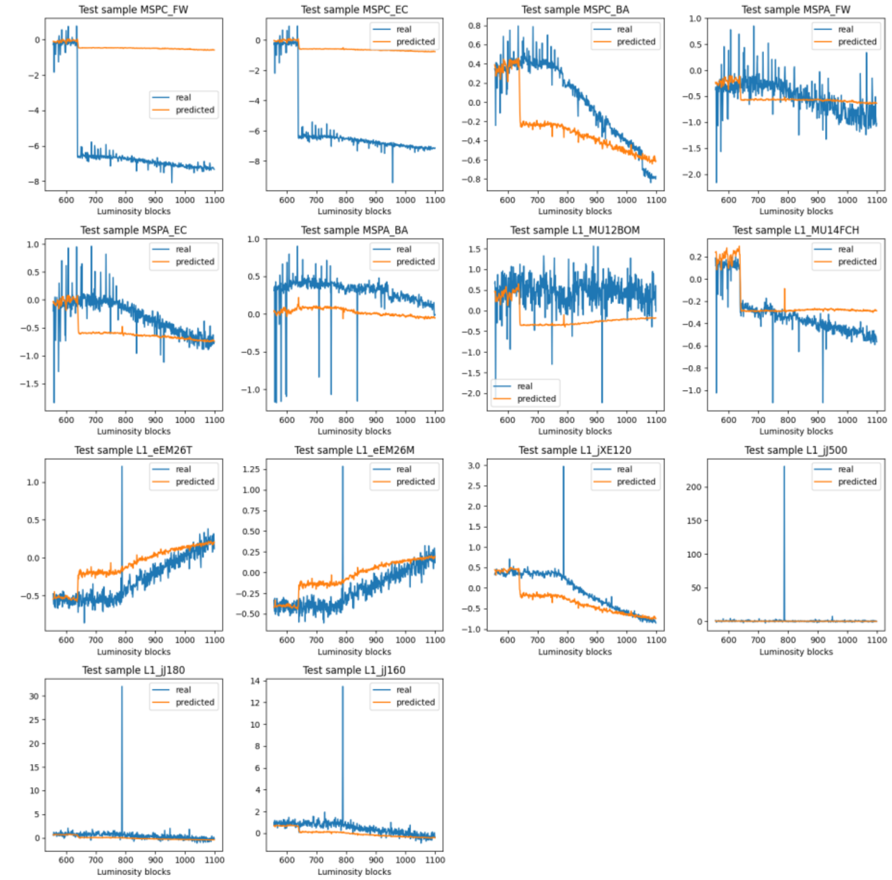


Model performance – muon end cap shutdown

In early June, a sector of the muon end cap was disabled (1/8 of one side of detector). We used this test data to see how our model responded. Can see a clear spike in MSE when disabled corresponding to the large variation between real and predicted values.

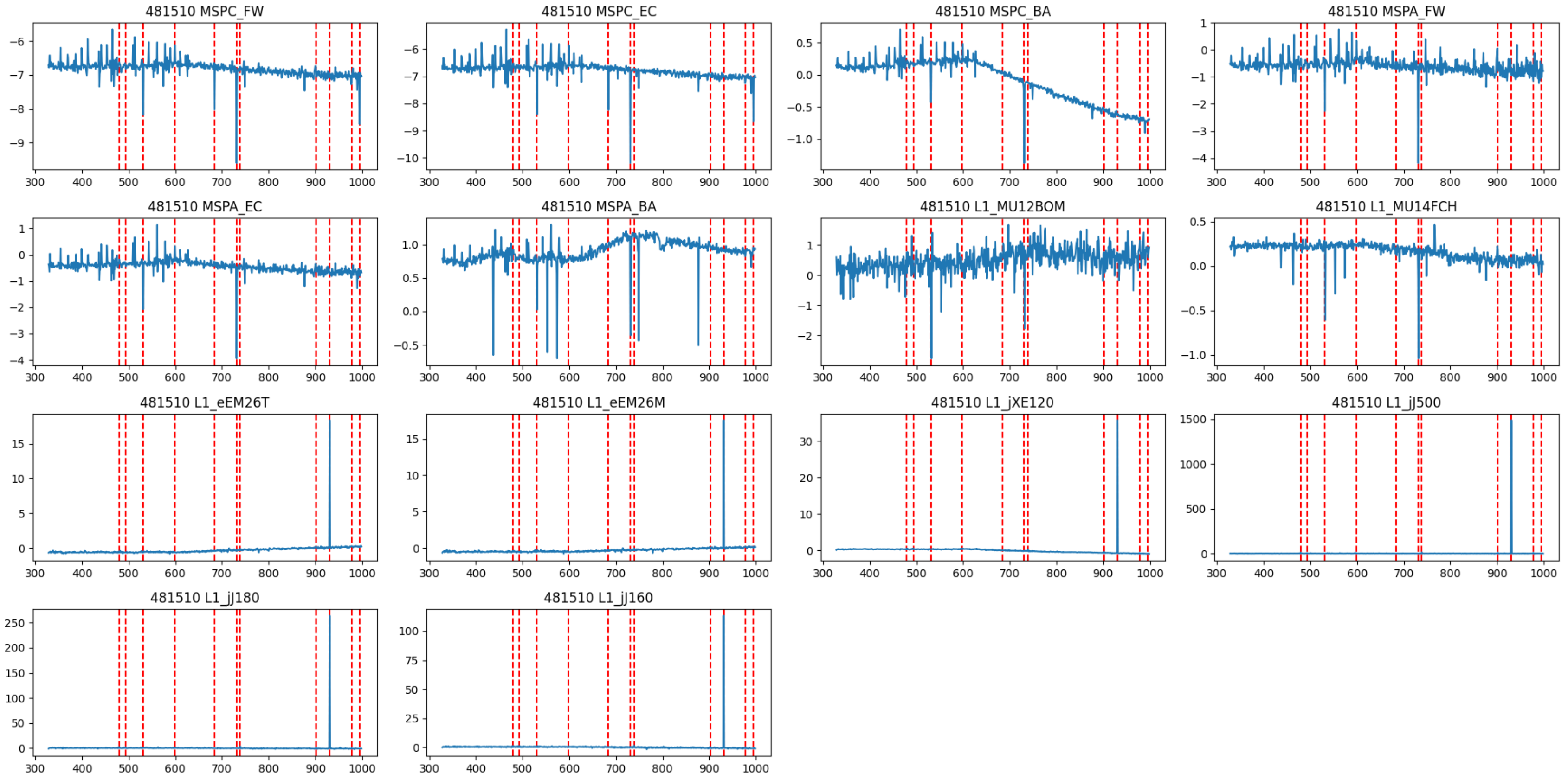






Model performance – anomalies in recent runs

- Using a recent run as test data, we check what our model identifies as anomalies
- With threshold = 8, model flags 11 anomalies in run 481510



Next Steps

- Separate models for constant pileup and decaying pileup or separate models for different classes of features
- Optimizing for better performance in low pileup region
- Piecewise renormalization
- Scaling up for more extensive feature set
- Continuous learning
- Structure for online usage

Acknowledgements:

Thank you to my supervisors Mario Campanelli and Antoine Marzin for their guidance this summer. I'd also like to acknowledge the University of Michigan REU program for their support.