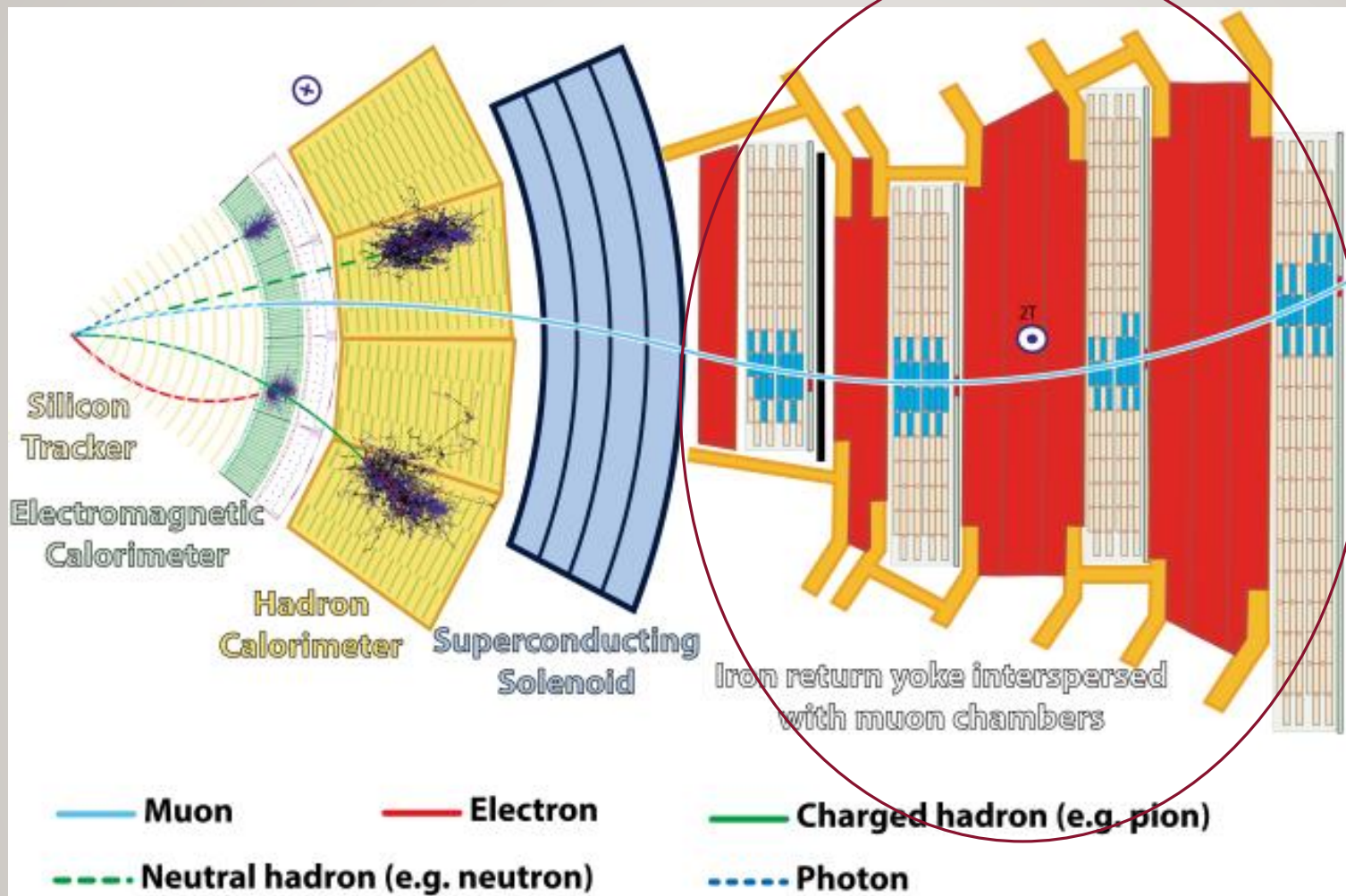


Anomaly Detection using Machine Learning for Data Quality Monitoring in the CMS Experiment

Presenter: Agrima Seth

Supervisors: Gianluca Cerminara
Adrian Alan Pol

THE CMS DETECTOR



Focus Area:

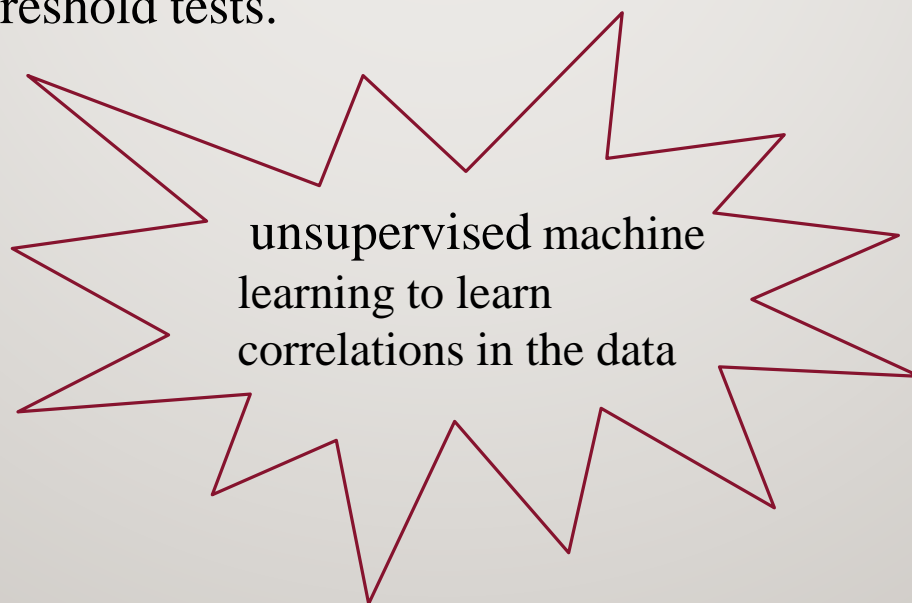
- Test Muon Hit counts in single electronic channels
- Identify Anomalous regions (aka chamber)

Current System

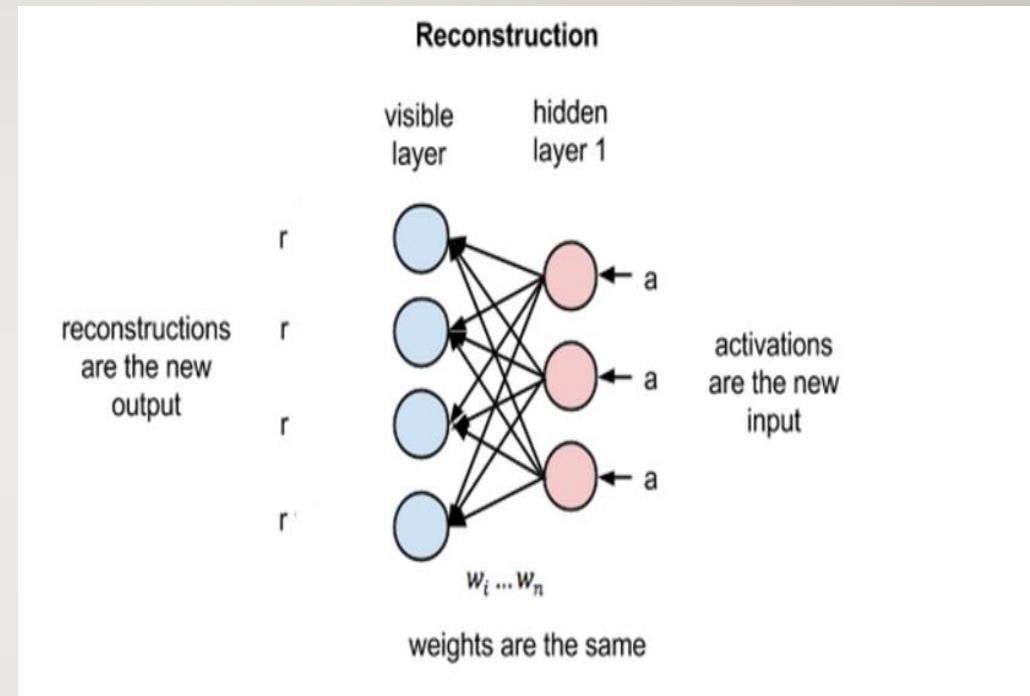
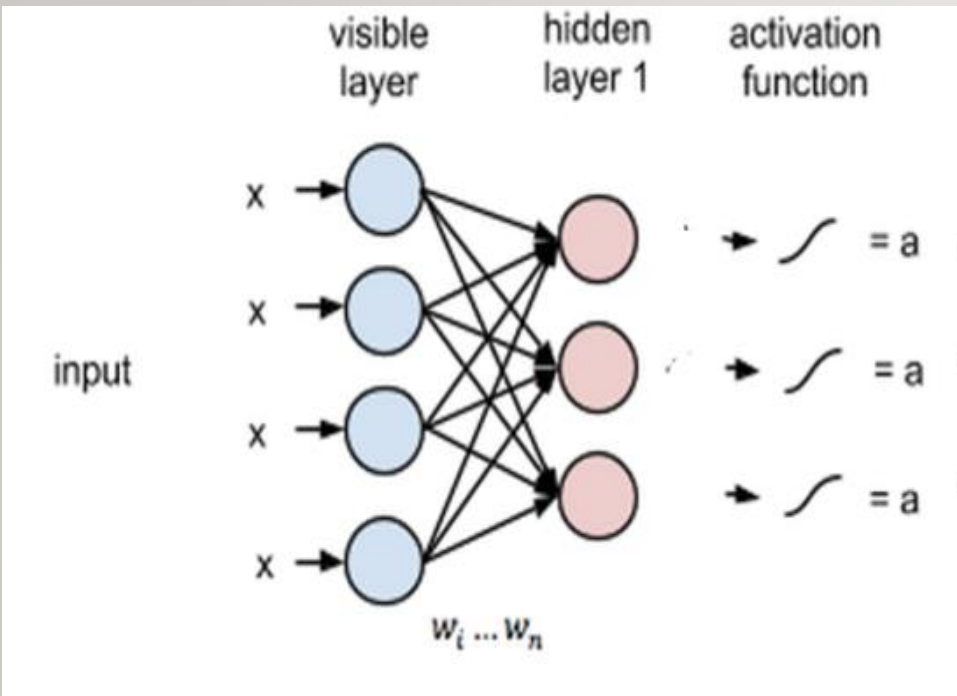
- The data quality assessment is based on plots displayed by the GUI
- Human inspection is the key to identifying anomalies
- Decisions are based on guidelines set by experts.

Potential to Enhance Current Framework

- **SIZE** : Overwhelming data (order of 100 million read out channels); making it difficult to monitor each quantity.
- **HUMAN INSPECTION**: Decisions will vary per person.
 - A machine learning model will generate reproducible results.
- **APPLIED THRESHOLDS**: The statistics tests often look only for expected "features"
 - Go beyond fixed threshold tests.



Autoencoders



(Dau, Hoang Anh, Vic Ciesielski, and Andy Song. "Anomaly detection using replicator neural networks trained on examples of one class." *Asia-Pacific Conference on Simulated Evolution and Learning*. Springer, Cham, 2014.)

Our Model

DATASET A

- Only good runs (size= 5990)
- Training (80%), Testing (20%)

DATASET B

- Mix healthy and known anomalous runs.(size = 4000)
- Testing(100%)

DATASET C

- 10 artificial anomalies
- Testing (100%)

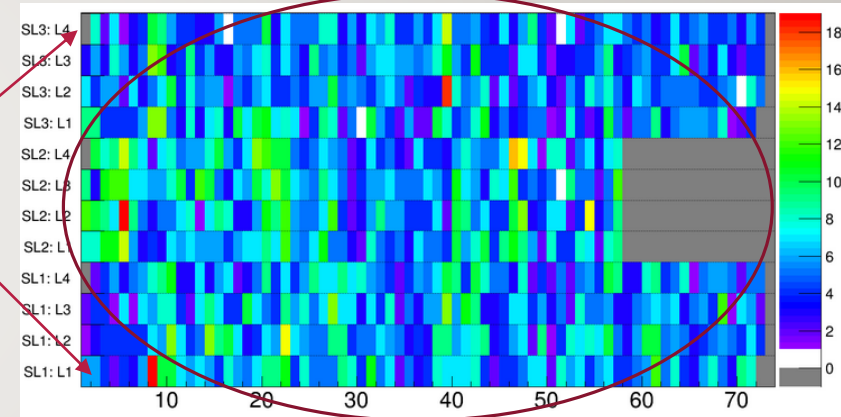
Features:

Number : 15

Comparison between: Topology, Median per layer(12 layers)
&
Topology, Mean per layer(12 layers)

Position in detector

Chamber



Data Preprocessing:

Robust Normalization (removes the median and scales the data according to the quantile range) or Min-Max normalization to spot relative differences between different chambers

Activation Functions Used:

- Encoder layer: Relu
- Decoder layer: Relu

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

What Next ?

SHORT TERM (next 1 week):

- Test performance different activation function.
- Characterize performance of various normalization techniques.

LONGER TERM:

- Enrich feature set (e.g. add moment of distribution)
- Look at different Unsupervised learning techniques

Stare at the detector and brainstorm



Agrima Seth
Computer Science Student



<https://www.linkedin.com/in/agrimaseth/>



agrima@seas.upenn.edu



<http://agrimaseth.github.io/>