Anomaly Detection in the LINAC

Davion Washington
North Carolina Agricultural and Technical State University | SIST

Jason St. John, Brian Schupbach
Accelerator Division
Fermi National Accelerator Laboratory

Abstract

Anomaly detection is a data analysis strategy that identifies data observations that diverge from the dataset's normal behavior. The anomalous data (or data in outage) in comparison to pre-anomalous (data prior) data can indicate the instance of the Linac approaching an anomaly phase. Using dimensional reduction to cluster 240 dimensions of data into two dimensions making it visible to compare similar clustering groups of anomalies.

Introduction

Fermilab houses a variety of particle accelerators such as the Linac. The Linac, or linear accelerator, is a straight accelerator that measures 500 ft. It brings proton beams up to energies of 400 MeV, providing proton beam for the Booster accelerator and the rest of the chain of accelerators

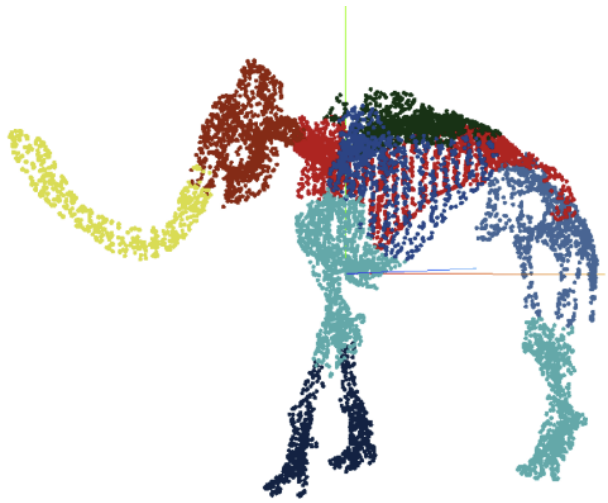
Objective

The objectives of this project included using clustering of down-projected machine state space as anomaly detection to predict beam outages and to identify classes of anomalous, unplanned outages in the Linac (Linear Accelerator). Beam outages mean less beam for research, and lost time, while identifying outage classes is a necessary first step to automating mitigation or prevention measures.

To fit a UMAP [1] (Uniform Manifold Approximation and Projection) transform to 240-dimensional data to obtain a 2D state space showing the clusters of known beam-outage states and apply the fixed transform to pre-onset data to test predictive capabilities.

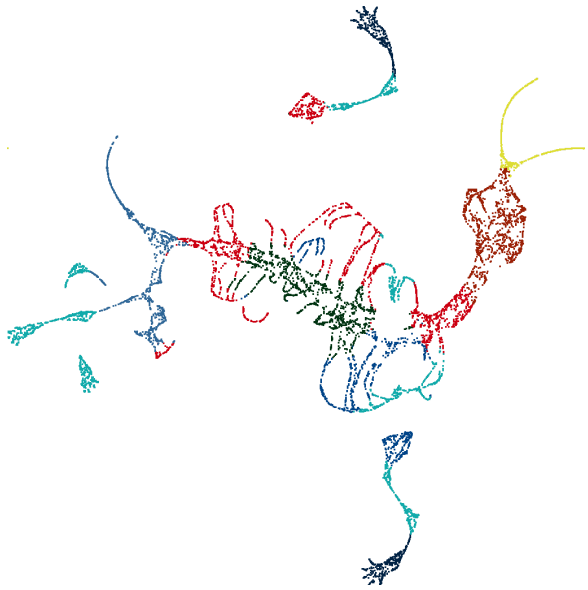
UMAP

UMAP is an algorithm for dimension reduction based on manifold learning techniques and ideas from topological data analysis[1]. It allows data to be clustered by similarity from hundreds of dimensions to an easily visualized two, for example dimensionality reduction is a powerful tool for machine learning practitioners to visualize and understand large, high dimensional datasets.



(Figure 1.1). UMAP original 3D data

The data before the dimensional reduction of UMAP (Figure 1) [2]. Data analysis of 3 dimensional data is difficult to visualize how far or close the data may be in terms of the next cluster and detecting anomalies.



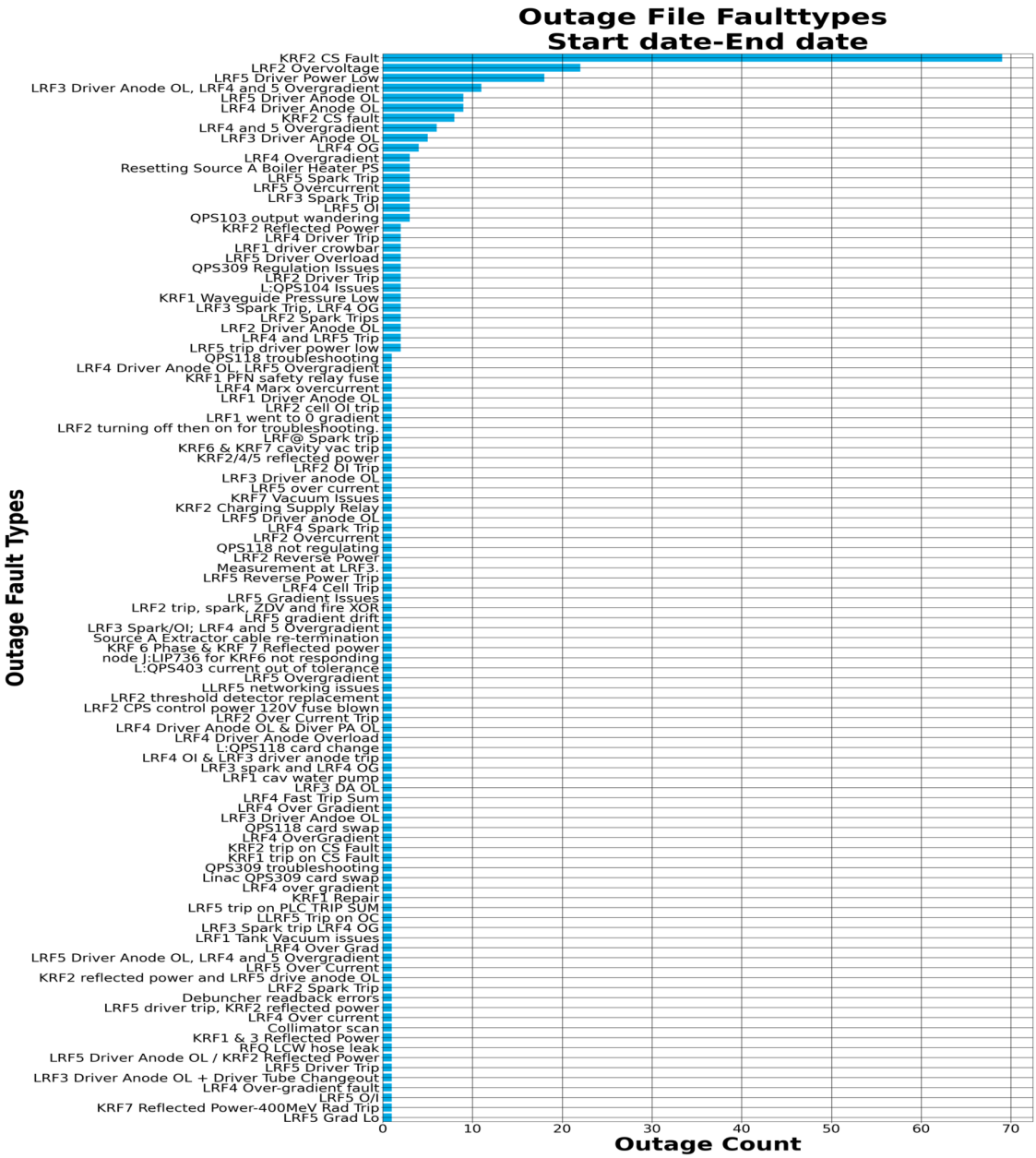
(Figure 1.2). UMAP 2D projection

After the clustering of UMAP (Fig 1.2), a 2 dimensional representation of the data shown above (Fig. 1.1) is presented showing the data flattened and indicating local structure with the corresponding colors. This machine learning tool helps visualize high-dimensional data in lower dimensional space while still allowing clustering analysis to work.

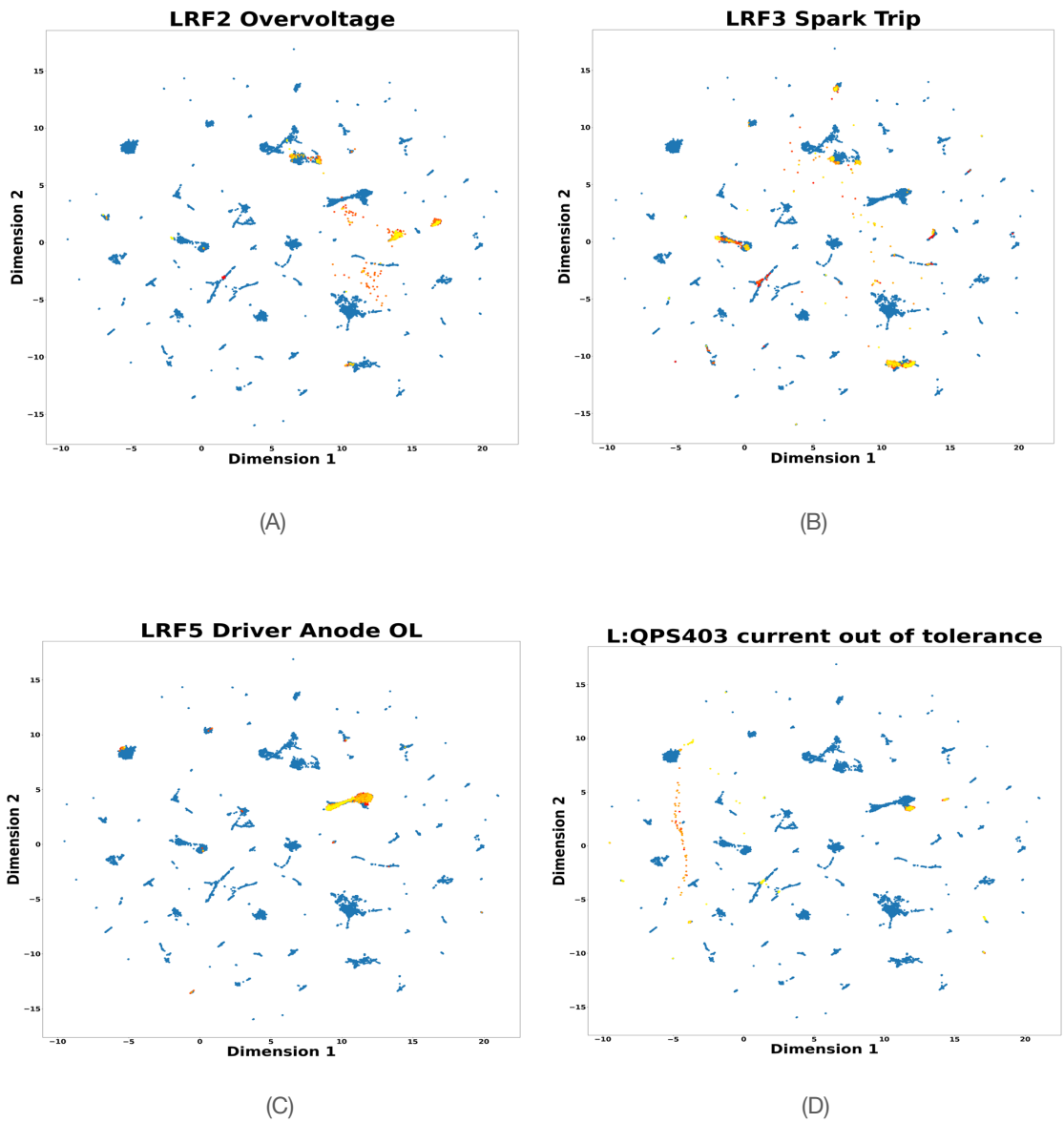
Results

I was able to apply a fixed UMAP transform to data from 24 Linac devices in files of 30 minutes of pre-anomaly data in 10 second intervals. Then I successfully was able to overlay anomalous

data with the pre-anomaly data in 2 dimensions to show possible predictions of upcoming beam-outages in the Linac. My research found that for certain fault types, all 30 minutes of pre-anomaly data clustered in a characteristic way. For some faults, many instances of faults were available, but it is not true for most faults. (See Fig. 2) Where we can judge what is typical, repeating patterns suggest predictive ability for those specific faults.



(Figure 2). Fault Type Histogram



(Figure 3). Anomalous data (Blue) overlayed with pre-anomaly data (Red to yellow). Labeled by developing anomaly.

The anomalous data is all of the files while in an outage,while the pre-anomaly data is the 30 minutes leading up to an outage in each corresponding outage file displayed in each title. In Fig. 3 B & C, the pre-anomaly data shows a local structure similar to the anomalous, hinting at an upcoming anomaly. However, in Fig. 3 A & D, the pre-anomaly data does not overlay a designated cluster matching the anomalous data.

Conclusion

Some pre-anomaly data formed recognizable patterns that resembled beam-outage states for one observed repeated fault, “LRF5 Driver Anode OL,”which occurred 10+ times. These results could convey that pre-anomaly data often shows similar activity of an upcoming anomaly a short time prior to when the anomaly period begins. While other pre-anomaly data formed

instances outside of the anomalous data clusters, such as “LRF2 Overvoltage” and “L:QPS403 current out of tolerance.” This could be explained if the pre-outage data was collected accidentally (without a real outage at its end), or that the 30 minute time frame is too long or too short, or maybe the projection was pushed past its limits of validity. This gives a lot of room for future work to dive further in the analysis of the Linac outage files.

Acknowledgment

I would like to thank Jason St. John, the L-CAPE team, Brillina Wang, and Aleksandar Dyulgerov. I also appreciate Fermilab and the L-CAPE team for presenting me with this research opportunity.

References

[1] McInnes, L, Healy, J, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, ArXiv e-prints 1802.03426, 2018

[2]Andy Coenen, Adam Pearce. “Understanding UMAP”, https://pair-code.github.io/understanding-umap/