# FAULT CLASSIFICATION USING MACHINE LEARNING

By:
Jyothi Yadav
Mihika Dhariwal
Adrija Singh

# PROBLEM STATEMENT

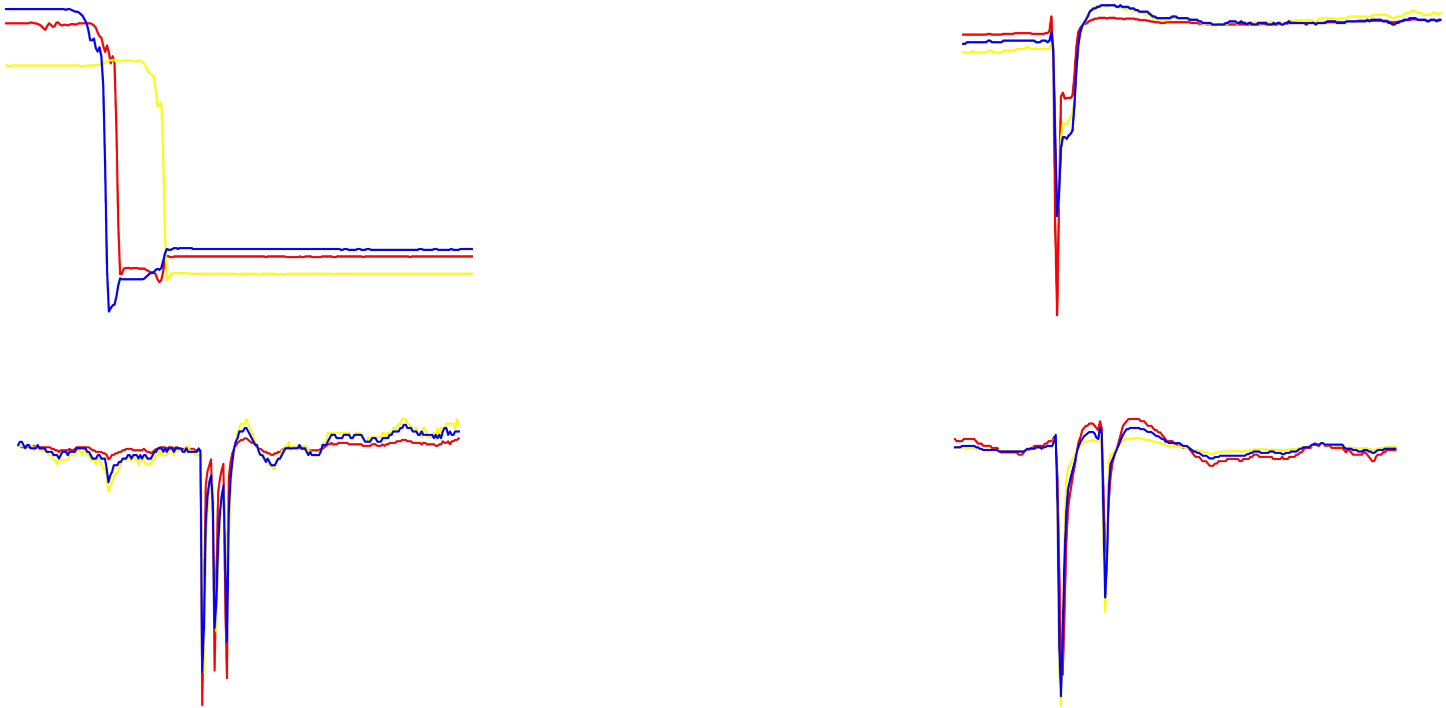
- The aim of our project is to use machine learning to detect and classify faults.

- Faults can be classified into the following categories: line to ground (RG, YG, BG), line to line (RY, YB, RB), double line to ground (RYG, YBG, RBG), triple line (RYB), triple line to ground (RYBG) and faultless.

- Different types of faults within a system may not have consistent fault signatures across all occurrences. Instead, a single type of fault can exhibit multiple fault signatures, which can vary based on the specific event or conditions leading to the fault. For example, events like tripping or auto-reclosure can have different time sequence graphs even though they cause the same type of fault, making it challenging to effectively classify them

- The primary objective in this context is to develop an effective methodology for identifying and classifying these diverse fault signatures associated with different types of faults.

# DIFFERENT FAULT SIGNATURES



# CLUSTERING ON HEATMAPS

- We used the VGG deep learning neural network to extract features from heatmaps, which represent our data.
- To simplify our analysis and reduce computational complexity, we applied Principal Component Analysis (PCA) to reduce the dimensionality of the feature vectors.
- The reduced feature vectors were then fed into the K-means clustering algorithm, which grouped similar data points into clusters.
- We determined the optimal number of clusters (k) using an elbow graph.

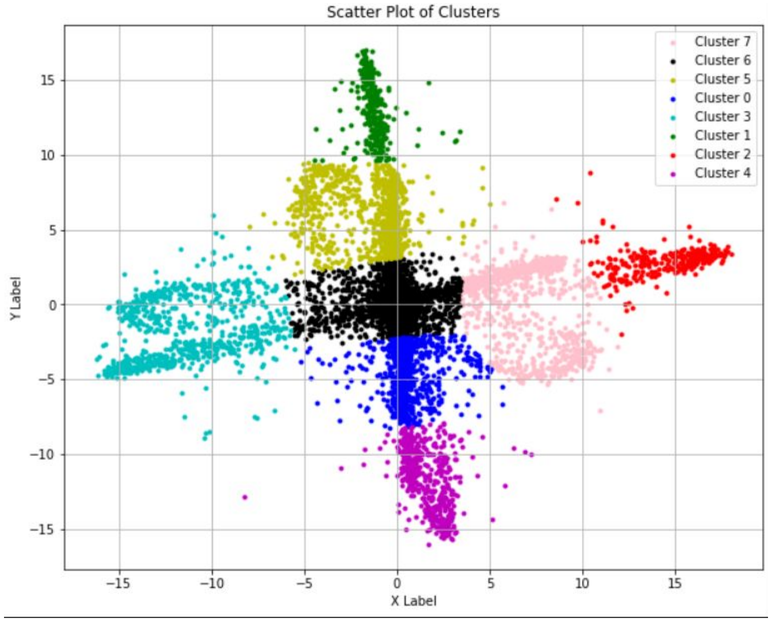
# CLUSTERING BASED ON TIME -SEQUENCE GRAPHS

- We generated time sequence graphs from our raw signal data and used them to perform feature extraction for clustering purposes.
- We hypothesize that this approach will be more effective in capturing important signal characteristics such as edges, voltage sags, and fault signatures compared to heatmaps.
- We base this hypothesis on the assumption that the VGG model, a deep convolutional neural network architecture, is better suited for identifying patterns and features in sequential data, which aligns with the nature of time sequence graphs.
- After extracting relevant features from the time sequence graphs using the VGG model, we apply K-means clustering, to group similar signals together based on their feature representations.
- Results grouped signals with tripping event into a single cluster, however other clusters obtained did not effectively group other type of fault events.

# CLUSTERING BASED ON DETAILED WAVELET COEFFICIENTS

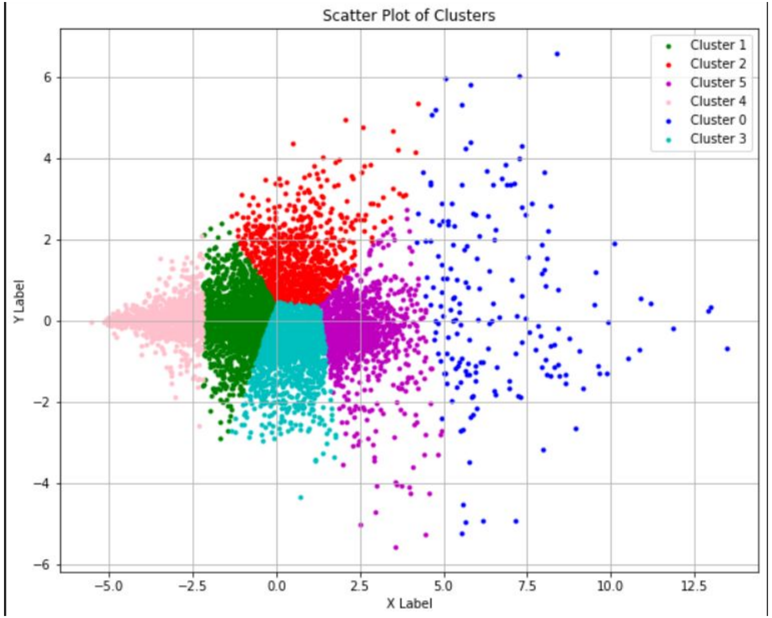
- Discrete wavelet transform was performed on each phase voltage of the time series data.
- Basis: Wavelet coefficients capture certain time-frequency characteristics unique to each type of fault.
- Different wavelet types (Haar, Db1)  with different decomposition levels (1,2) were used.
- The set of detailed coefficients obtained for each phase were concatenated into a single feature vector and clustering was performed.
- Alternative Approach: The norm of detailed coefficient of each phase voltage was calculated, and concatenated into a single feature vector, and clustering was performed.
- Result: Clustering was performed on detailed coefficients, but perhaps on another property of the time series data unknown to us, and not different types of faults as expected.

# RESULTS




Wavelet: Haar, level: 2

All detailed coefficients considered to generate the feature vector




Wavelet: Db1, level: 1

Norm of detailed coefficients considered to generate feature vector

# SUPERVISED LEARNING

- Heatmap images were generated by applying wavelet transform on raw signal data.
- These heatmaps were trained on a CNN to identify 12 different types of faults: RG, YG, BG, RY, YB, RB, RYG, YBG, RBG, RYB, RYBG, Faultless
- Result: Training accuracy: 90% Testing accuracy: 85%
- Aim: To increase the accuracy of the model while minimizing overfitting

Sample Heatmap

# FUTURE SCOPE: USING COHERENCE

- Idea: Using the coherence function to identify the portions of the signal that deviate significantly from a reference signal, which may be considered fault-free.
- Isolating these potentially faulty segments, to reduce the amount of data needed to be analyzed, making the subsequent fourier or wavelet transforms to be applied on it, more focused and computationally efficient.
- Performing clustering on these coefficients to identify groups which have same type of fault signatures.