# Fault Detection and Classification in 3-phase Electrical Power Transmission Line

Ziyang Yuan, Hai Xi, Yanbo Peng, Wentao Jiang
University of Rochester
May 7, 2024

# Overview

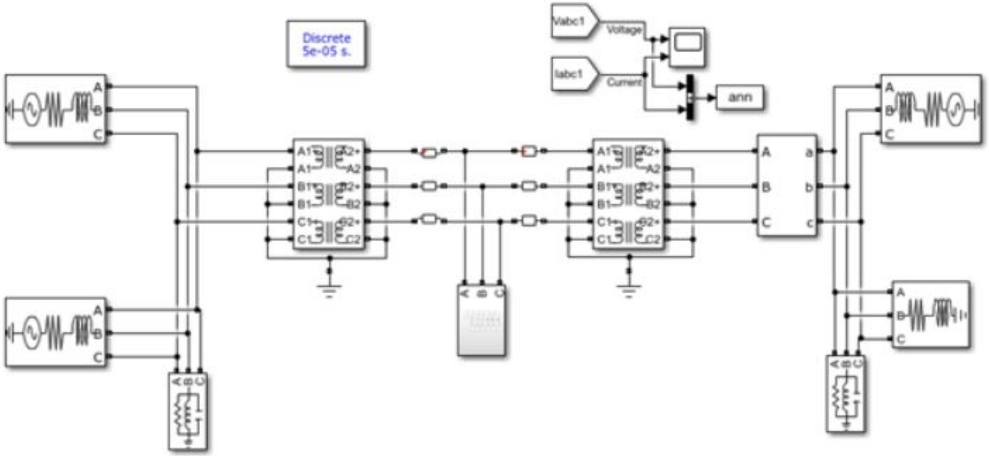
1. Project Background
2. Model Selection & Evaluation
3. Tackling Low Accuracy in Complex Faults
4. Model Performance
5. Real World Applications

# 1. Background

# Background

## Purposes

- Electrical faults–stability and reliability.
- Quick detection and classification of these faults
- Input: Voltages and Currents; Output: Fault types
- Develop a real-time electrical fault detection & classification system, i.e. using the data of voltages and current to monitor the conditions of the transmission lines




Simulink simulation

# Dataset

Objective:
Develop a real-time electrical fault detection and classification system

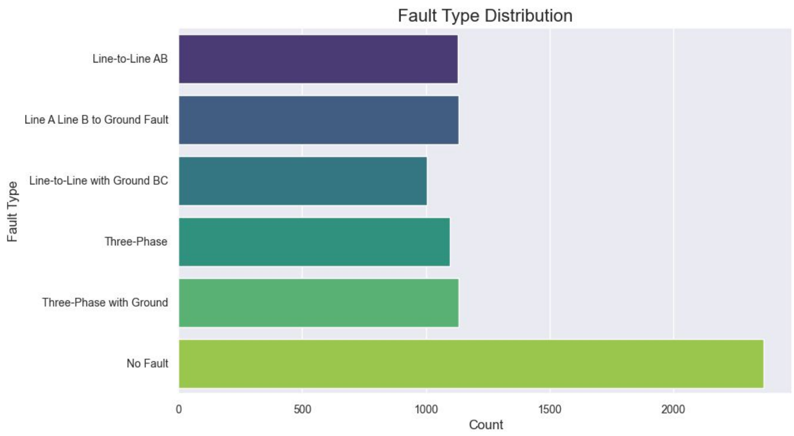
| | G | C | B | A | Ia | Ib | Ic | Va | Vb | Vc |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 0 | 0 | 1 | -151.291812 | -9.677452 | 85.800162 | 0.400750 | -0.132935 | -0.267815 |
| **1** | 1 | 0 | 0 | 1 | -336.186183 | -76.283262 | 18.328897 | 0.312732 | -0.123633 | -0.189099 |
| **2** | 1 | 0 | 0 | 1 | -502.891583 | -174.648023 | -80.924663 | 0.265728 | -0.114301 | -0.151428 |
| **3** | 1 | 0 | 0 | 1 | -593.941905 | -217.703359 | -124.891924 | 0.235511 | -0.104940 | -0.130570 |
| **4** | 1 | 0 | 0 | 1 | -643.663617 | -224.159427 | -132.282815 | 0.209537 | -0.095554 | -0.113983 |

Input:
- Current in Line A (Ia)
- Current in Line B (Ib)
- Current in Line C (Ic)
- Voltage in Line A (Va)
- Voltage in Line B (Vb)
- Voltage in Line C (Vc)

Outputs:
- '0000': 'No Fault',
- '1000': 'Single Line to Ground A',
- '0100': 'Single Line to Ground B',
- '0010': 'Single Line to Ground C',
- '0011': 'Line-to-Line BC',
- '0101': 'Line-to-Line AC',
- '1001': 'Line-to-Line AB',
- '1010': 'Line-to-Line with Ground AB',
- '0101': 'Line-to-Line with Ground AC',
- '0110': 'Line-to-Line with Ground BC',
- '0111': 'Three-Phase',
- '1111': 'Three-Phase with Ground',
- '1011': 'Line A Line B to Ground Fault'

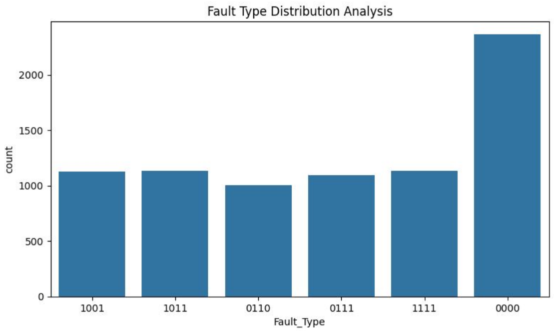
# Models Selection & Evaluation

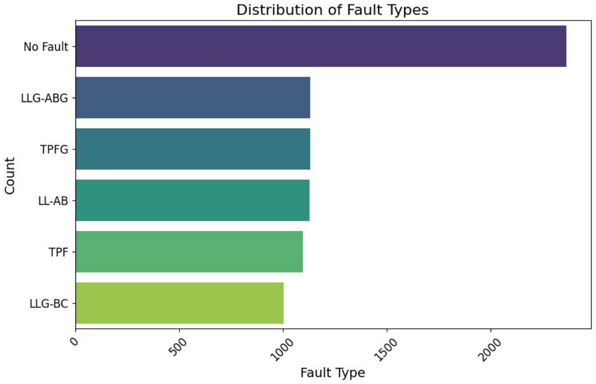
# Model Selection

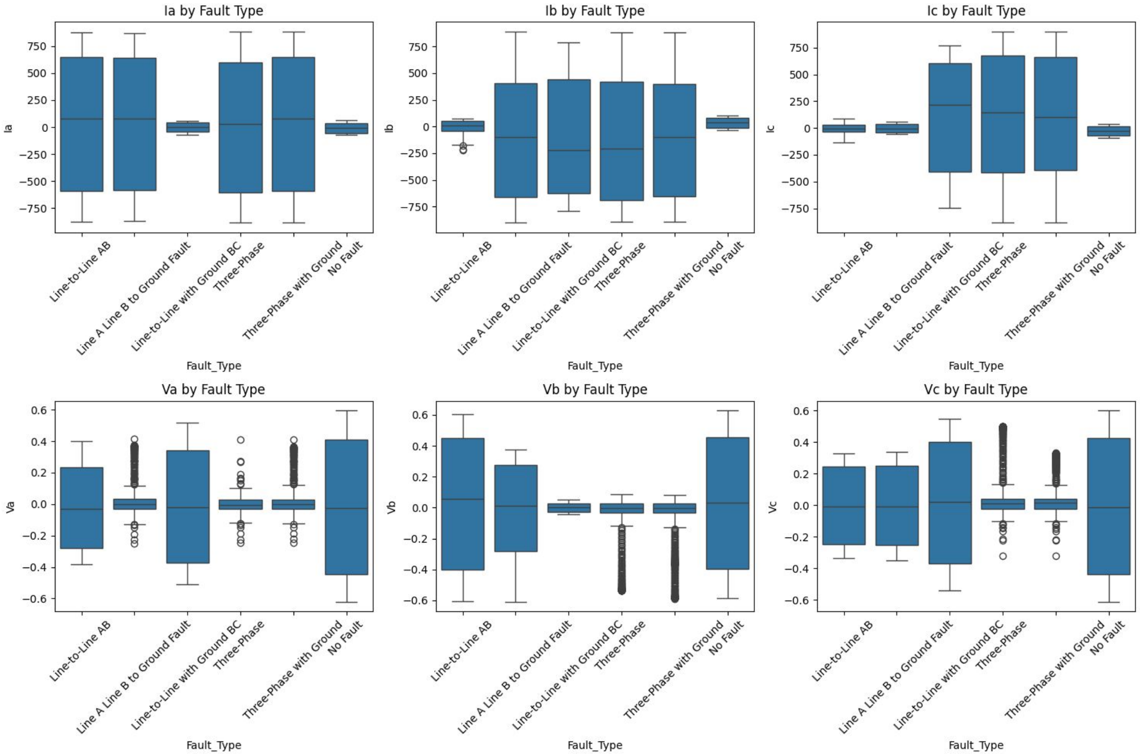



- A Classification Problem
  - Decision Tree Classifier
  - **Random Forest Classifier**
  - Support Vector Classifier (linear kernel)
  - K-Nearest Neighbor

# Feature Analysis



# Feature Engineering

- **Standardization** and **Encoding** for Model Input
    - Min-Max Scaler
    - Label Encoder
- Composite Features
    - PolynomialFeatures (Next)

$$\left(X_{Scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}\right)$$, range [0,1]

$$(B, R, G) \rightarrow (2, 0, 1)$$

# Polynomial Feature Transformation

**Examples**:

- Interaction

  $(I_a V_a)$

- Higher Degree

  $(I_a{}^3)$

- **Interaction Terms and Higher Degree Terms**
- **Modeling the interactions** between different variables
- Pattern and Dimensionality
  - n original features with polynomial degree of d
  - $\binom{n+d}{d}$
- **Limitations**
  - Scalability and Overfitting (use d=2)

# Hyperparameter Tuning

- Handling of Imbalanced Data
  - Synthetic Minority Over-sampling Technique (SMOTE)
- Number of Trees
- Features of Trees
  - "Gini" impurity
  - Min sample leaf = 1
  - Min sample split = 2

$$\text{Gini} = 1 - \sum_{i=1}^{n} p_i^2$$

# Performance

Performance Metrics for Random Forest:
Accuracy: 88.30%
Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| Line A Line B to Ground Fault | 1.00 | 1.00 | 1.00 | 227 |
| Line-to-Line AB | 1.00 | 1.00 | 1.00 | 226 |
| Line-to-Line with Ground BC | 1.00 | 1.00 | 1.00 | 201 |
| No Fault | 1.00 | 1.00 | 1.00 | 473 |
| Three-Phase | 0.58 | 0.61 | 0.59 | 219 |
| Three-Phase with Ground | 0.60 | 0.57 | 0.58 | 227 |
| accuracy | | | 0.88 | 1573 |
| macro avg | 0.86 | 0.86 | 0.86 | 1573 |
| weighted avg | 0.88 | 0.88 | 0.88 | 1573 |

Confusion Matrix:

```
 [[227   0   0   0   0   0]
 [  0 226   0   0   0   0]
 [  0   0 201   0   0   0]
 [  0   0   0 473   0   0]
 [  0   0   0   0 133  86]
 [  0   0   0   0  98 129]]
```

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1 Score} = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}$$

Support = Frequency of Actual Class

# Tackling Low Accuracy in Complex Faults

# Integrating Electrical Engineering Insights

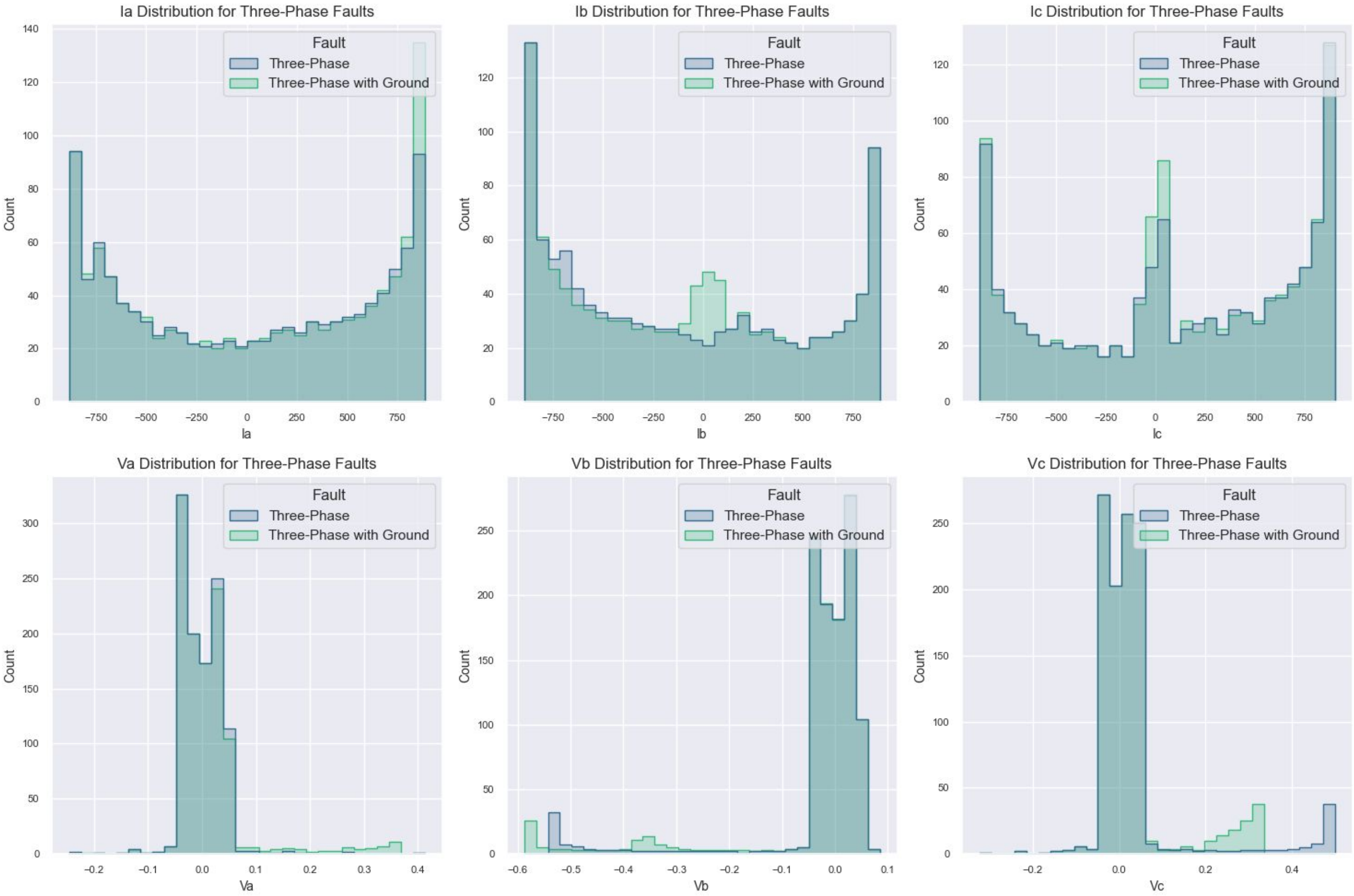
- **Identifying Weak Points**
  - **Handle imbalanced data**
  - 'Three-Phase'
  - 'Three-Phase with Ground'
- **Domain Knowledge**
  - Adding new features

```
Performance Metrics for Random Forest:
Accuracy: 88.30%
Classification Report:
                               precision    recall  f1-score   support

Line A Line B to Ground Fault       1.00      1.00      1.00       227
              Line-to-Line AB       1.00      1.00      1.00       226
  Line-to-Line with Ground BC       1.00      1.00      1.00       201
                     No Fault       1.00      1.00      1.00       473
                  Three-Phase       0.58      0.61      0.59       219
      Three-Phase with Ground       0.60      0.57      0.58       227

                     accuracy                           0.88      1573
                    macro avg       0.86      0.86      0.86      1573
                 weighted avg       0.88      0.88      0.88      1573

Confusion Matrix:
 [[227   0   0   0   0   0]
 [  0 226   0   0   0   0]
 [  0   0 201   0   0   0]
 [  0   0   0 473   0   0]
 [  0   0   0   0 133  86]
 [  0   0   0   0  98 129]]
```


Ia Distribution for Three-Phase Faults
Ib Distribution for Three-Phase Faults
Ic Distribution for Three-Phase Faults
Va Distribution for Three-Phase Faults
Vb Distribution for Three-Phase Faults
Vc Distribution for Three-Phase Faults
Fault
Three-Phase
Three-Phase with Ground
Count
Ia
Ib
Ic
Va
Vb
Vc

# Domain Knowledge

1. **Zero Sequence Component for Current And Voltage**:
   - Current: Average of phases A, B, and C.
   - Voltage: Average of phases A, B, and C.

```
# Zero Sequence Components for Current and Voltage
poly_features_df['ZeroSeqCurrent'] = (poly_features_df['Ia'] + poly_features_df['Ib'] + poly_features_df['Ic']) / 3
poly_features_df['ZeroSeqVoltage'] = (poly_features_df['Va'] + poly_features_df['Vb'] + poly_features_df['Vc']) / 3
```

# Domain Knowledge

1. **Zero Sequence Component for Current And Voltage**:
   - Current: Average of phases A, B, and C.
   - Voltage: Average of phases A, B, and C.
2. **Phase Angle Difference**:
   - Approximated by the product of the currents and voltages of all three phases

```
# Phase Angle Differences (approximated by product of current and voltage)
poly_features_df['PhaseAngleDiffI'] = poly_features_df['Ia'] * poly_features_df['Ib'] * poly_features_df['Ic']
poly_features_df['PhaseAngleDiffV'] = poly_features_df['Va'] * poly_features_df['Vb'] * poly_features_df['Vc']
```

# Domain Knowledge

1. **Zero Sequence Component for Current And Voltage**:
   - Current: Average of phases A, B, and C.
   - Voltage: Average of phases A, B, and C.
2. **Phase Angle Difference**:
   - Approximated by the product of the currents and voltages of all three phases
3. **Total Harmonic Distortion**:
   - Approximated for each phase to measure waveform distortion

```
# Total Harmonic Distortion (THD) - Approximation
poly_features_df['THD_Ia'] = np.sqrt(poly_features_df['Ia^2'] - poly_features_df['Ia']**2) / poly_features_df['Ia']
poly_features_df['THD_Ib'] = np.sqrt(poly_features_df['Ib^2'] - poly_features_df['Ib']**2) / poly_features_df['Ib']
poly_features_df['THD_Ic'] = np.sqrt(poly_features_df['Ic^2'] - poly_features_df['Ic']**2) / poly_features_df['Ic']
```

# Domain Knowledge

1. Zero Sequence Component for Current And Voltage:

```
# Voltage and Current Ratios
poly_features_df['V_I_Ratio_A'] = poly_features_df['Va'] / poly_features_df['Ia']
poly_features_df['V_I_Ratio_B'] = poly_features_df['Vb'] / poly_features_df['Ib']
poly_features_df['V_I_Ratio_C'] = poly_features_df['Vc'] / poly_features_df['Ic']
```

- Approximated by the product of the currents and voltages of all three phases

3. **Total Harmonic Distortion**:

- Approximated for each phase to measure waveform distortion

4. **Voltage and Current Ratios**:

- Ratio of voltage to current for each phase, indicating impedance characteristics

# Expanding Feature Set

- **Old Features**
  - Current and voltage measurements for each phase
    - (Ia, Ib, Ic, Va, Vb, Vc)
  - Square and cross-product terms( From FE)
    - (Ia^2, Ia Ib, Va Vb...)

- **New Features Added**

  Incorporated advanced features derived from domain expertise to capture more complex and subtle phenomena associated with faults
  - **Zero Sequence Components**
  - **Phase Angle Differences**
  - **Total Harmonic Distortion** (THD)
  - **Voltage and Current Ratios**

# Model Performance

```
Performance Metrics for Random Forest:
Accuracy: 99.94%
Classification Report:
                                  precision    recall  f1-score   support

Line A Line B to Ground Fault        1.00      1.00      1.00       227
             Line-to-Line AB         1.00      1.00      1.00       226
  Line-to-Line with Ground BC        1.00      1.00      1.00       201
                    No Fault         1.00      1.00      1.00       473
                 Three-Phase         1.00      1.00      1.00       219
     Three-Phase with Ground         1.00      1.00      1.00       227

                    accuracy                             1.00      1573
                   macro avg         1.00      1.00      1.00      1573
                weighted avg         1.00      1.00      1.00      1573

Confusion Matrix:
 [[227   0   0   0   0   0]
 [  0 226   0   0   0   0]
 [  0   0 201   0   0   0]
 [  0   0   0 473   0   0]
 [  0   0   0   0 219   0]
 [  0   0   0   0   1 226]]
```

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{F1 Score} = 2 \cdot \frac{Precision \times Recall}{Precision + Recall}$$

Support = Frequency of Actual Class

# Performance and Real-World Applications

- Significant **Performance Improvements**

**From 65% to 100%. Overall from 87% to 99.34%**

- Industrial **Application**

**Transmission lines; Electrical Motors; Stoves; HVAC …**

# Thank you!