Supplementary Material:
A Machine Learning Based Approach for Comprehensive Fault Diagnosis in Transmission Lines

Computers & Electrical Engineering, vol. XX, no. XX, pp. XX–XX, 202X.

Isternândia Araújo França, Cynthia Wanick Vieira, Daniel Correa Ramos, Lara Hoffmann Sathler,

Eduardo G. Carrano

Dear editor and reviewers,

We moved some contents of the manuscript to a Supplementary Material in order to comply with the 25 page limit defined by the journal. If the paper is accepted, this Supplementary Material will be made available through GitHub. Its URL will be properly cited into the document.

1 Introduction

This supplementary material companions the manuscript entitled *A Machine Learning Based Approach for Comprehensive Fault Diagnosis in Transmission Lines*, published on Computers & Electrical Engineering, vol. XX, no. XX, pp. XX–XX, year 202X. It contains detailed material regarding the features considered by the classifiers and some additional results. It is important to emphasize that it is not intended to be a self contained document. Such a supplementary material should be read jointly with the original manuscript.

2 List of Symbols

- **LG**; **LL**; **LLG**; **LLL** : line-to-ground; line-to-line; line-to-line-to-ground; line-to-line-to-line
- **LI**; **MI**; **HI** : low impedance; medium impedance; high impedance
- **C1**; **C2**; **C3**; **C4**; **C5** : fire under line; lightning strikes; wind storms/bird droppings; short circuit through vegetation; pollution on insulators
- **S**; **SE**; **N/NE**; **CW** : south region; southeast region; north/northeast regions; center-west region
- **Dec-Apr**; **May-Jul**; **Aug-Nov**: December to April; May to July; August to November
- P : positive sequence
- N : negative sequence
- Z : zero sequence
- T_{Fault} : the fault instant estimated
- T_{Window} : phasor evaluation time window (set by the user)
- median(R) : median of sample R
- MAD(R) : Mean Absolute Deviation of sample R
- $F_{i,j}$: value of instance j for feature i
- $F_{i,j}^{st}$: standardized value of instance j for feature i
- $A(\mathcal{C})$: accuracy of classifier $\mathcal{C}$
- T_P; T_N; F_P; F_N : number of: true positives; true negatives; false positives; false negatives
- Z_F : fault impedance
- V_K^{Bef}; V_K^{Aft} : voltage phasors before and after the fault
- I_K^{Bef}; I_K^{Aft} : current phasors before and after the fault
- real(Y) : real component of the complex number Y

3 List of Acronyms

- **ANN** : Artificial Neural Networks
- **AUC** : Area Under Curve
- **BPNN** : Back Propagation Based Network
- **BTW2** : Second order Butterworth filter

- **CART** : Classification and Regression Tree
- **CNN** : Convolution Neural Network
- **CSHT** : Chi-Square Hypothesis Test
- **DT** : Decision Tree
- **DWT** : Discrete Wavelet Transform
- **ELM** : Extreme Learning Machines
- **FDOST** : Fast Discrete Orthogonal S-Transform
- **FT** : Fourier Transform
- **GPS** : Global Positioning System
- **LSE** : Least Squares Estimation
- **MFQL** : Modified fuzzy-Q-learning
- **ML** : Machine Learning
- **MLP** : Multilayer Perceptron
- **MODP** : Magnitude of Differential Power
- **PMU** : Phasor Measurement Unit
- **PNN** : Probabilistic Neural Network
- **RBF** : Radial Basis Function
- **RFE** : Recursive Feature Elimination
- **RMNN** : Rough Membership Neural Network
- **ROC** :Receiver Operating Characteristic
- **RTDS** : Real Time Digital Simulator
- **SNN** : Single-Nearest-Neighbor classifier
- **SVM** : Support Vector Machines
- **TW** : Traveling Wave Line protection relay

4 Features

The features considered by phase, ground, and impedance level classifiers are divided in two main classes: basic features and composite features. The basic features are the ones extracted directly from relay oscillographies. On the other hand, the composite features are the ones that are evaluated based on basic features. The composite features were conceived taking into account the nature of each classification problem, in order to allow the use of lower complexity classifiers. A list of the eight basic features considered and the 40 composite features proposed in this work are described along this section.

4.1 Basic Features

Eight basic features are extracted from pre- and post-fault phasors, as follows:

- **BF01** – Magnitude of voltage phasor before fault: $\left|V_K^{Bef}\right|$
- **BF02** – Angle of voltage phasor before fault: $\angle V_K^{Bef}$
- **BF03** – Magnitude of voltage phasor after fault: $\left|V_K^{Aft}\right|$
- **BF04** – Angle of voltage phasor after fault: $\angle V_K^{Aft}$
- **BF05** – Magnitude of current phasor before fault: $\left|I_K^{Bef}\right|$
- **BF06** – Angle of current phasor before fault: $\angle I_K^{Bef}$
- **BF07** – Magnitude of current phasor after fault: $\left|I_K^{Aft}\right|$
- **BF08** – Angle of current phasor after fault: $\angle I_K^{Aft}$

In **BF01** to **BF08**, K is the phase under analysis:

- Phase classifier: $K \in \{A,B,C\}$ (A, B, and C are the system phases).
- Impedance level classifier: $K \in \{P,N,Z\}$ (P, N, and Z are the positive sequence, negative sequence, and zero sequence of symmetric components).
- Ground classifier: $K \in \{Z\}$.

4.2 Composite Features

The composite features are evaluated based on the basic ones. The set of composite features considered for the phase classification is the following:

- **CFP01** – Absolute phase voltage magnitude variation:
  $\left|\frac{V_K^{Aft}}{V_K^{Bef}}\right|$

- **CFP02** – Absolute phase current magnitude variation:
$$\left|\frac{I_K^{Aft}}{I_K^{Bef}}\right|$$

- **CFP03** – Relative phase voltage magnitude before fault:
$$\frac{\left|V_K^{Bef}\right|}{\frac{1}{3}\left(\left|V_A^{Bef}\right|+\left|V_B^{Bef}\right|+\left|V_C^{Bef}\right|\right)}$$

- **CFP04** – Relative phase current magnitude before fault:
$$\frac{\left|I_K^{Bef}\right|}{\frac{1}{3}\left(\left|I_A^{Bef}\right|+\left|I_B^{Bef}\right|+\left|I_C^{Bef}\right|\right)}$$

- **CFP05** – Relative phase voltage magnitude after fault:
$$\frac{\left|V_K^{Aft}\right|}{\frac{1}{3}\left(\left|V_A^{Aft}\right|+\left|V_B^{Aft}\right|+\left|V_C^{Aft}\right|\right)}$$

- **CFP06** – Relative phase current magnitude after fault:
$$\frac{\left|I_K^{Aft}\right|}{\frac{1}{3}\left(\left|I_A^{Aft}\right|+\left|I_B^{Aft}\right|+\left|I_C^{Aft}\right|\right)}$$

- **CFP07** – Relative phase voltage magnitude difference:
$$\frac{\left|V_K^{Aft}-V_K^{Bef}\right|}{\frac{1}{3}\left(\left|V_A^{Aft}-V_A^{Bef}\right|+\left|V_B^{Aft}-V_B^{Bef}\right|+\left|V_C^{Aft}-V_C^{Bef}\right|\right)}$$

- **CFP08** – Relative phase current magnitude difference:
$$\frac{\left|I_K^{Aft}-I_K^{Bef}\right|}{\frac{1}{3}\left(\left|I_A^{Aft}-I_A^{Bef}\right|+\left|I_B^{Aft}-I_B^{Bef}\right|+\left|I_C^{Aft}-I_C^{Bef}\right|\right)}$$

- **CFP09** – Relative phase voltage magnitude variation:
$$\frac{\left|\frac{V_K^{Aft}}{V_K^{Bef}}\right|}{\frac{1}{3}\left(\left|\frac{V_A^{Aft}}{V_A^{Bef}}\right|+\left|\frac{V_B^{Aft}}{V_B^{Bef}}\right|+\left|\frac{V_C^{Aft}}{V_C^{Bef}}\right|\right)}$$

- **CFP10** – Relative phase current magnitude variation:
$$\frac{\left|\frac{I_K^{Aft}}{I_K^{Bef}}\right|}{\frac{1}{3}\left(\left|\frac{I_A^{Aft}}{I_A^{Bef}}\right|+\left|\frac{I_B^{Aft}}{I_B^{Bef}}\right|+\left|\frac{I_C^{Aft}}{I_C^{Bef}}\right|\right)}$$

- **CFP11** – Phase current angle difference:
$$\angle I_K^{Aft}-\angle I_K^{Bef}$$

- **CFP12** – Phase impedance angle difference:
$$\angle\frac{V_K^{Aft}}{I_K^{Aft}}-\angle\frac{V_K^{Bef}}{I_K^{Bef}}$$

In **CFP01** to **CFP12**, $K \in \{A,B,C\}$.

Four composite features are calculated for ground classification:

- **CFG01** – Zero-sequence voltage magnitude difference:
  $\left|V_Z^{Aft} - V_Z^{Bef}\right|$
- **CFG02** – Zero-sequence current magnitude difference:
  $\left|I_Z^{Aft} - I_Z^{Bef}\right|$
- **CFG03** – Absolute zero-sequence voltage magnitude variation:
  $\left|\frac{V_Z^{Aft}}{V_Z^{Bef}}\right|$
- **CFG04** – Absolute zero-sequence current magnitude variation:
  $\left|\frac{I_Z^{Aft}}{I_Z^{Bef}}\right|$

Finally, 24 composite features are employed to classify impedance level (eight features per each symmetrical component sequence):

- **CFZ01-CFZ03** – K-sequence voltage magnitude variation:
  $\left|\frac{V_K^{Aft}}{V_K^{Bef}}\right|$
- **CFZ04-CFZ06** – K-sequence current magnitude variation:
  $\left|\frac{I_K^{Aft}}{I_K^{Bef}}\right|$
- **CFZ07-CFZ09** – K-sequence voltage magnitude difference:
  $\left|V_K^{Aft} - V_K^{Bef}\right|$
- **CFZ10-CFZ12** – K-sequence current magnitude difference:
  $\left|I_K^{Aft} - I_K^{Bef}\right|$
- **CFZ13-CFZ15** – K-sequence resistance before fault:
  $\mathrm{real}\left(\frac{V_K^{Bef}}{I_K^{Bef}}\right)$
- **CFZ16-CFZ18** – K-sequence resistance after fault:
  $\mathrm{real}\left(\frac{V_K^{Aft}}{I_K^{Aft}}\right)$

- **CFZ19-CFZ21** – K-sequence impedance variation:
$\left|\frac{\frac{V_K^{Aft}}{I_K^{Aft}}}{\frac{V_K^{Bef}}{I_K^{Bef}}}\right|$

- **CFZ22-CFZ24** – K-sequence resistance difference:
$\mathrm{real}\left(\frac{V_K^{Aft}}{I_K^{Aft}} - \frac{V_K^{Bef}}{I_K^{Bef}}\right)$

In **CFZ01** to **CFZ24**, $K \in \{P,N,Z\}$.

5 Additional Results

The confusion arrays obtained for round 1 of results are shown in Tables 1, 2, and 3. The ROC curves for the same data sets can be seen in Figure 1. These results are in agreement with the ones obtained for the combined data sets:

- Phase and ground classification performances were always close to 100%.
- The impedance level classifiers reached accuracy ratios around 90%.
- The performance for MI was always the worst among the three impedance level classes.

Table 1: Round 1 – Confusion arrays - Phase classifier

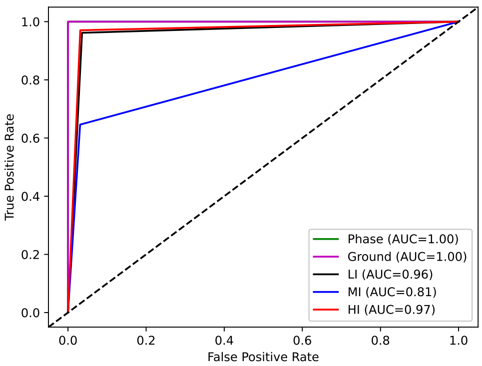
	SF:TL1		**SF:TL2**		**SF:TL3**		**SF:TL4**	
	P	**N**	**P**	**N**	**P**	**N**	**P**	**N**
P	2,357	0	2,382	0	2,387	0	2,436	13
N	0	3,643	0	3,618	0	3,613	10	3,541

Table 2: Round 1 – Confusion arrays - Ground classifier

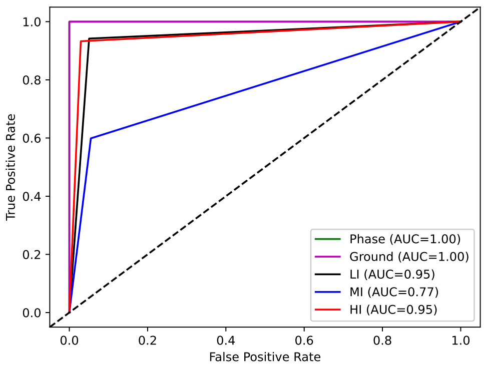
	SF:TL1		**SF:TL2**		**SF:TL3**		**SF:TL4**	
	P	**N**	**P**	**N**	**P**	**N**	**P**	**N**
P	610	0	611	0	633	0	586	0
N	0	589	0	587	0	558	0	595

Table 3: Round 1 – Confusion arrays - Impedance level

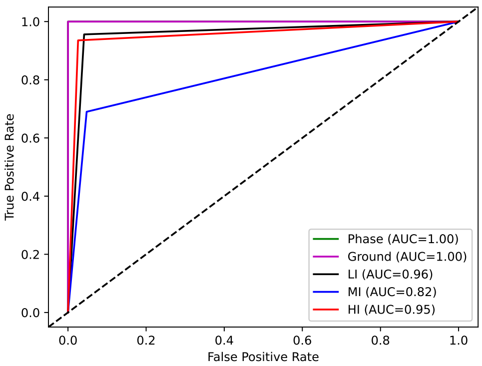
	SF:TL1			**SF:TL2**			**SF:TL3**			**SF:TL4**		
	LI	**MI**	**HI**	**LI**	**MI**	**HI**	**LI**	**MI**	**HI**	**LI**	**MI**	**HI**
LI	855	34	0	858	48	6	847	35	4	814	59	14
MI	35	126	34	47	110	25	38	138	24	50	113	49
HI	5	22	889	8	53	845	8	51	855	22	41	838



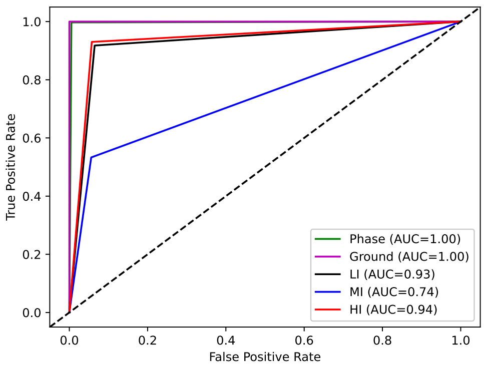
(a) RF:TL1 (phase and ground curves are superimposed)



(b) RF:TL2 (phase and ground curves are superimposed)



(c) RF:TL3 (phase and ground curves are superimposed)



(d) RF:TL4 (phase and ground curves are superimposed)

Figure 1: Round 1 – ROC curves and AUC values