Model Card: C100 Fault Identification Model, v1.2

Model Details

- Developed by researchers at Thomas Jefferson National Accelerator Facility, 2020
- Random Forest Classifier, v1.2

  {'bootstrap': False, 'class_weight': None, 'criterion': 'gini', 'max_depth': None, 'max_features': 'auto', 'max_leaf_nodes': None, 'min_impurity_decrease': 0.0, 'min_impurity_split': None, 'min_samples_leaf': 1, 'min_samples_split': 2, 'min_weight_fraction_leaf': 0.0, 'n_estimators': 800, 'n_jobs': None, 'oob_score': False, 'random_state': 11, 'verbose': 0, 'warm_start': False}
- Related Information: C. Tennant, A. Carpenter, T. Powers, A. Shabalina, L. Vidyaratne, K. Iftekharuddin "Superconducting Radio-Frequency Cavity Fault Classification Using Machine Learning at Jefferson Laboratory", Phys. Rev. Accel. Beams (*in preparation*).
- Contact: tennant@jlab.org, lasithav@jlab.org

Intended Use

- The model is intended for use in conjunction with data collected from a specially designed data acquisition system for C100-type superconducting radio-frequency (SRF) cavities at the Continuous Electron Beam Accelerator Facility (CEBAF). After a C100 cavity fault, the model uses waveform data recorded from all 8 cavities in the cryomodule to identify the type of fault.
- Primary users are CEBAF control room operators and SRF system experts.
- The model is only intended to work with the C100 cavities in CEBAF (excluding cryomodule 0L04) and to classify faults that occur during beam operations. Since the training data included faults where one or more individual cavities were bypassed, we believe this model should operate effectively in the presence of bypassed/disabled cavities.

Training Data

- The data was split into a train (70%) and test (30%) set with stratification to ensure that the train and test sets have approximately the same percentage of samples of each target class as the complete set.
- The data [DOI: 10.14462/MLFaultClassifier/1616675] is available at: https://github.com/JeffersonLab/AI_SRF_operations/tree/master/datasets/C100-2020-04-30
- The training set is comprised of a 1,662 × 192 feature matrix and associated labels. The labels have 8 possible classes and can take values of (Controls_Fault, E_Quench, Heat_Riser_Choke, Microphonics, Multi_Cav_Turn_Off, Quench_100ms, Quench_3ms, Single_Cav_Turn_Off).

Evaluation Data

- The data was split into a train (70%) and test (30%) set with stratification to ensure that the train and test sets have approximately the same percentage of samples of each target class as the complete set.
- The data [DOI: 10.14462/MLFaultClassifier/1616675] is available at: https://github.com/JeffersonLab/AI_SRF_operations/tree/master/datasets/C100-2020-04-30
- The testing set is comprised of a 713 × 192 feature matrix and associated labels. The labels have 8 possible classes and can take values of (Controls_Fault, E_Quench, Heat_Riser_Choke, Microphonics, Multi_Cav_Turn_Off, Quench_100ms, Quench_3ms, Single_Cav_Turn_Off).

Metrics

- Performance metrics for the fault identification model are summarized by the testing and cross-validation accuracies (Table 1) and the classification report (Table 2).

TABLE 1: Cross-validation and accuracy scores for the fault identification model.

10-fold cross-validation (%)	85.52 ± 3.65
accuracy (test data) (%)	87.66

TABLE 2: Classification report for fault identification model.

Fault Type	Precision	Recall	F1-score	Support
Controls_Fault	0.77	0.66	0.71	76
E_Quench	0.97	0.94	0.96	68
Heat_Riser_Choke	0.88	0.88	0.88	25
Microphonics	0.93	0.91	0.92	130
Multi_Cav_Turn_Off	0.85	0.84	0.84	61
Quench_100ms	0.92	0.91	0.91	100
Quench_3ms	0.82	0.85	0.84	66
Single_Cav_Turn_Off	0.85	0.93	0.89	187

Caveats and Recommendations

- Identifying the correct number and type of fault classes represented in the data is an area of current research.