Machine Learning Based Cavity Fault Classification and Prediction

Project ID: 2020-LDRD-LD2003

A. Shabalina, A. Carpenter, M. Rahman, C. Tennant, L. Vidyaratne

Introduction and Motivation

We report on the development of machine learning models for classifying C100 superconducting radio frequency (SRF) cavity faults in the Continuous Electron Beam Accelerator Facility (CEBAF) at Jefferson Lab [1]. CEBAF is a continuous-wave recirculating linac utilizing 418 SRF cavities to accelerate electrons up to 12 GeV through 5-passes. Of these, 96 cavities (12 cryomodules) are designed with a digital low-level RF system configured such that a cavity fault triggers waveform recordings of 17 RF signals for each of the eight cavities in the cryomodule. Subject matter experts (SME) can analyze the collected time-series data, identify which of the eight cavities faulted first, and classify the type of fault. This information is used to find trends and strategically deploy mitigations to problematic cryomodules. However, manually labeling the data is laborious and time-consuming. By leveraging machine learning, near real-time – rather than post-mortem – identification of the offending cavity and classification of the fault type has been implemented. We discuss the performance of the ML models during a recent physics run.

Machine Learning Model Development

To train models, we consider a dataset comprised of faults recorded during CEBAF beam operations between January 18, 2019, and March 9, 2020. The dataset contains 2,375 labeled, C100 cavity fault events. Each event is comprised of 17 time-series signals per cavity for each of the eight cavities in the cryomodule [2].

The primary challenge for machine learning applications utilizing time-series data is feature extraction, in which statistical parameters (or features) are computed from the raw data signals. These features serve as an intermediate representation of the data and are used as model inputs. For our data, six autoregressive coefficients are computed for each signal.

Prior to training, we split the labeled events 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. We withhold the test data as the unseen data that we use for the final model evaluations. A variety of classification models from the scikit-learn [3] library were trained, including k-Nearest Neighbors, Decision Tree, Support Vector, and Gaussian Naive Bayes. We also included ensemble models such as the Bagging Classifier, Random Forest, Extra Trees, and Gradient Boosting. For each model, we

use *k*-fold cross-validation to estimate the performance, ensuring the test data remains untouched. Figure 1 displays the results of training models for fault classification. The accuracy plotted on the vertical axis refers to the ability of the models to reproduce labels given in the training set data.

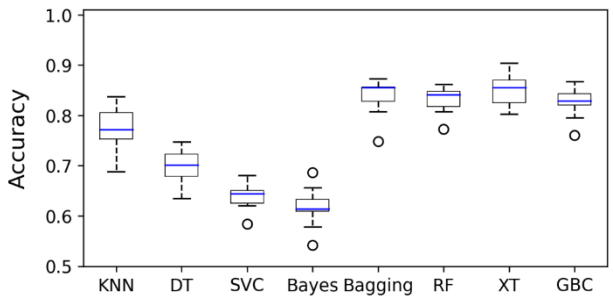

FIG. 1: Boxplots showing accuracy scores from a 10-fold cross-validation analysis of several algorithms for cavity fault classification. The blue line denotes the median, the box spans the interquartile range (IQR), the upper (lower) whiskers indicate values 1.5×IQR above (below) the upper (lower) box boundary, with data beyond the whiskers represented as open markers. Ensemble models (four rightmost boxplots) exhibit the best performance.

The Random Forest Classifier was chosen to model both the cavity identification and fault classification because of its good performance but also for its robustness against overfitting. The next step is to optimize the model through hyperparameter tuning. After optimization, the 10-fold cross-validation score (based on training data) is compared with predictions from the withheld test set. Table 1 summarizes the results and indicates that overfitting, and the inability of a model to generalize learning to data outside the training set, is avoided.

TABLE 1: Cross-validation and accuracy scores for each Random Forest model. The accuracy scores were generated by applying the model on the withheld test data set.

	Cavity Identification	**Fault Classification**
10-fold cross-validation (%)	87.97 ± 1.81	85.52 ± 3.65
Accuracy (test data) (%)	87.94	87.66

Results

We deployed fully functional machine learning models to the CEBAF's control system software environment on March 10. Following a C100 cavity fault event, the software automatically collects RF signals and saves the data to disk. The machine learning models use this data to label the cavity and fault type, and save their outputs to a database. This information is presented to operators via a web-based RF waveform viewer (see Fig. 2), and to SME through routine email-based reports. Thus far, the operational models have labeled 4,217 fault events, and have informed a number of mitigations.

We analyzed the 312 labeled fault event occurring in March, prior to the COVID shutdown, for model performance against an independent SME labeling. The models identified the first cavity to fault and the type of fault with accuracies of 84.9% and 78.2%, respectively. Though the performance is lower than during testing, the effectiveness of the system does not depend on the ability to accurately classify a single, standalone event. Rather, it is used to look for trends over time.

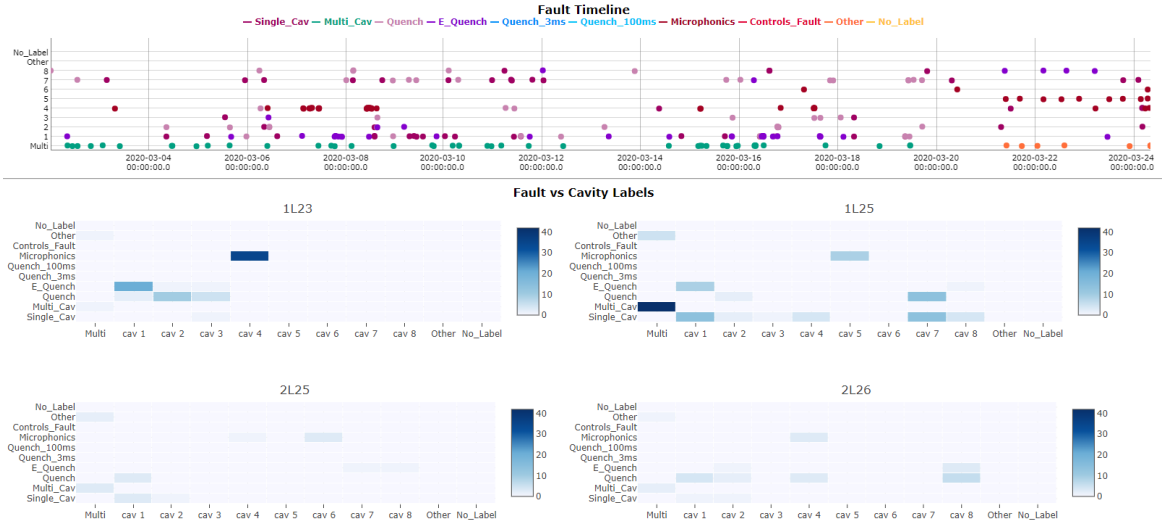

FIG. 2: Online interface to display results of the ML model outputs. The top plot is a timeline that displays fault events where the markers are color-coded according to type and the cavity identifier is represented by its placement on the vertical axis. Heat maps capture the spatial, rather than temporal distribution, of fault events for each cryomodule.

Deep Learning Model Development

We also investigated replacing current machine learning models with their deep learning counterparts. Deep learning is a sub-field of machine learning that is based on learning successive layers of increasingly meaningful representations of the data using a large scale artificial neural network (ANN) architectures. The primary advantage of methods based on learning data representations is that it avoids the extensive hand engineering of feature extraction and feature selection steps that require subject matter expertise. Rather than human developed feature extraction algorithms, the recorded data serves as examples and the model learns the necessary features for inference.

A variety of deep learning architectures were analyzed; including a recurrent neural network model based on long-short term memory (LSTM), a recurrent autoencoder (AE) [4], and two variations of a deep convolutional neural network (CNN) [5]. All deep learning architectures are developed to perform both cavity recognition and fault recognition tasks simultaneously. Fig 3 represents the deep bidirectional LSTM based cavity and fault classification architecture [4].

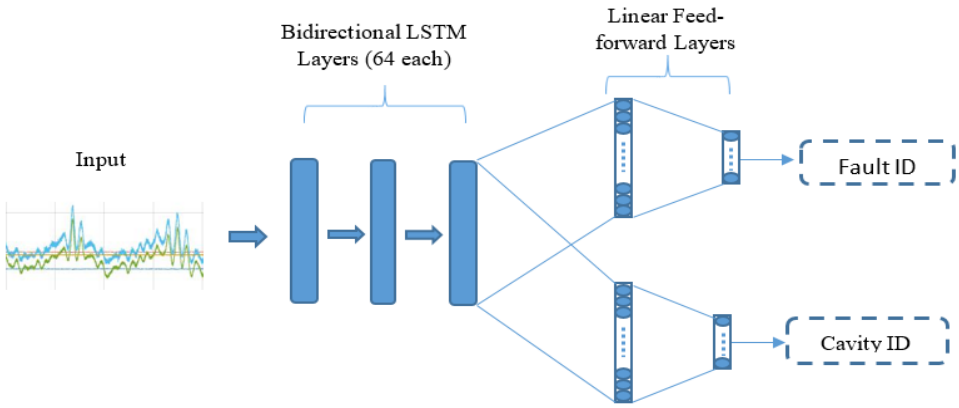

FIG. 3: Deep bidirectional LSTM with simultaneous cavity and fault classification architecture

The LSTM is a variety of ANNs that are specifically designed to learn temporal features in time series signals. Consequently, the deep LSTM architecture in Fig 3 uses raw RF signals as input to determine the failed cavity and fault type of the event. Fig 4 represents the deep CNN architecture developed for cavity and fault classification [5].

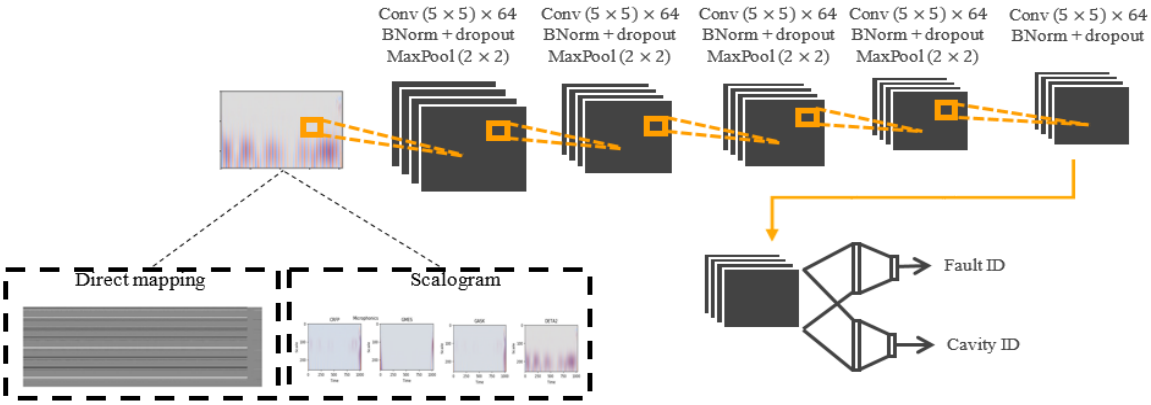

FIG. 4: Deep CNN with simultaneous cavity and fault classification architecture

CNNs are particularly designed to process image data and have shown to be quite successful at many image-processing tasks encountered in computer vision applications. The convolutional layers with trainable weight kernels allow CNNs to efficiently learn complex features that extend through the image space. As CNN are tailored to process images, we must convert the RF signals into image representations for compatibility. Consequently, we experiment with two image representation methods as shown in Fig 4. The image representations are subsequently processed by our CNN architecture to determine the failed cavity and fault type of the event.

Table 2 summarizes the performance of the above deep models obtained with the same dataset utilized for the machine learning models.

TABLE 2: Training, validation, and testing accuracy scores for each deep learning model. The testing accuracy scores were generated by applying the model on the withheld test data set.

DL Model	**Cavity Classification**			**Fault Classification**		
	Input Size	*Train/Validation Accuracy (%)*	*Test Accuracy (%)*	*Input Size*	*Train/Validation Accuracy (%)*	*Test Accuracy (%)*
Deep LSTM *(Raw Signals)*	136 × 1024	86.4% / 84.97%	84.6%	136 × 1024	77.1% / 73.4%	74.1%
CNN *(Direct Mapping)*	136 × 1024	86.4% / 84.97%	84.6%	136 × 1024	77.1% / 73.4%	74.1%
CNN *(Scalogram)*	32 × 256 × 1024	97.0% / 45.3%	45.0%	32 × 256 × 1024	97.4% / 59.7%	58.9%

Most deep learning models perform adequately for both tasks with over 80% accuracy for cavity classification and over 70% accuracy for fault classification. However, we observe that the DL model performance is consistently lower than that of the machine learning model (compare with table 1). This is in large part due to the relatively small training set. The expectation is that as more labeled data becomes available, the performance of these data-hungry deep learning models will improve.

Unsupervised Learning Analysis

The motivation for using unsupervised machine learning techniques in the context of the C100 cavity data is to identify fault types. At present, a SME determines the fault classes. Therefore, when presented with unfamiliar patterns in the data, it raises the question of whether it is a variant of a known fault type or if it represents something completely new. Invoking dimensionality reduction techniques and then applying clustering algorithms may provide a data-driven method for identifying the number of fault types represented in the data, and potentially reveal new fault types.

Initial results of applying unsupervised learning to the C100 cavity fault data provide interesting clues about the underlying nature of the fault types, raise additional questions, and motivate further investigation [6]. This is an important analysis, the results of which can be used as feedback to help SME and identify possible variants of a fault type (i.e. "microphonics type A", "microphonics type B") or identify a new fault altogether. Consequently, a more rigorous unsupervised learning technique based on a shapelet learning schemes currently being developed. Shapelets are essentially a category of learnable temporal features that describe important aspects of a given time series signal. The framework under development aims to learn the best shapelets that represent RF faults and perform unsupervised clustering simultaneously. The final shapelets obtained after the learning cycles will be used to cluster the input data into distinct categories and may help to differentiate between RF fault types present in the input data

Summary

We have described the development and implementation of machine learning and deep learning models at CEBAF to automate the task of classifying C100 SRF cavity faults. The ability to automatically label fault events immediately reduces the burden on SMEs and the time-consuming task of inspecting and labeling data manually. Having a labeled dataset is of interest at two different timescales. On the one hand, near real-time results after a fault event provide guidance to control room operators for appropriate responses (i.e. one type of fault might necessitate a reduction in cavity gradient, while another type of fault may indicate a larger problem and require intervention from the system experts). On the other hand, using the aggregate statistics and breaking down faults according to cryomodule and cavity provides data-driven guidance for maintenance and/or upgrade activities.

Work is continuing on a variety of fronts. We have deployed our machine learning models into CEBAF's control system environment with software tools that help inform decisions around C100 fault mitigations. Over 4000 fault events have been classified by the models over months of beam operations. We stand ready to update our models with an additional 4275 SME labeled fault events, and are working with our SRF SME to develop standard procedures for responding to trends reported by the operational machine learning models.

References

[1] C. Tennant et al., "Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory", Phys. Rev. Accel. Beams *(under review)*.

[2] https://github.com/JeffersonLab/AI_SRF_operations/tree/master/datasets/C100-2020-04-30
DOI 10.14462/MLFaultClassifier/1616675

[3] scikit-learn, https://scikit-learn.org/stable

[4] L. Vidyaratne, “Deep Recurrent Learning for C100 Cavity and Fault Classification”, Jefferson Lab Technical Note 20-027 (2020).

[5] L. Vidyaratne, “Convolutional Neural Networks for C100 Cavity and Fault Classification”, Jefferson Lab Technical Note 20-030 (2020).

[6] C. Tennant, “Using Unsupervised Learning to Analyze C100 Fault Type Data”, Jefferson Lab Technical Note 20-020 (2020).