2020

Superconducting Radio-Frequency Cavity Fault Classification Using Machine Learning at Jefferson Laboratory

Chris Tennant

Adam Carpenter

Tom Powers

Anna Shabalina Solopova

Lasitha Vidyaratne

See next page for additional authors

Authors

Chris Tennant, Adam Carpenter, Tom Powers, Anna Shabalina Solopova, Lasitha Vidyaratne, and Khan
Iftekharuddin

Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory

Chris Tennant⬤, Adam Carpenter, Tom Powers,
Anna Shabalina Solopova⬤, and Lasitha Vidyaratne
Jefferson Laboratory, Newport News, Virginia 23606, USA

Khan Iftekharuddin
Old Dominion University, Norfolk, Virginia 23529, USA

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. CEBAF is a continuous-wave recirculating linac utilizing 418 SRF cavities to accelerate electrons up to 12 GeV through five 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 are able to analyze the collected time-series data and 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 postmortem—identification of the offending cavity and classification of the fault type has been implemented. We discuss performance of the machine learning models during a recent physics run. Results show the cavity identification and fault classification models have accuracies of 84.9% and 78.2%, respectively.

I. INTRODUCTION

The Continuous Electron Beam Accelerator Facility (CEBAF) at Jefferson Laboratory is a continuous-wave recirculating linac (see Fig. 1) [1,2]. In September 2017, CEBAF completed an energy upgrade to extend its energy reach from 6 to 12 GeV. To meet this milestone, 88 newly designed cavities (representing 11 cryomodules) were installed. Each cryomodule (comprised of eight seven-cell cavities) is capable of 100 MV energy gain and is regulated with an associated digital low-level rf system (LLRF). These are known as the C100 cryomodules.

As a nuclear physics user facility, CEBAF aims to maximize beam availability to the four experimental stations. Currently, the largest contributor to short machine downtime trips (defined as events which are resolved in less than 5 min) are caused by superconducting radio-frequency (SRF) system faults. During 2019, there were an average of 4.1 rf trips per hour, which translated to 58 min of beam

time lost to rf recovery each day. The time lost in the experimental halls is effectively greater, since data are discarded 30 sec before the trip and 30 sec after recovery. The C100 modules, in particular, were responsible for nearly a quarter of the downtime due to short trips across all accelerator subsystems. Presently, the only operational parameter that operators can change in order to reduce the trip rate is the cavity gradient, which means that the trip rate is inversely related to CEBAF's energy reach. Ideally, sufficient gradient overhead would exist to meet the demands of both tolerable trip rates and desired energy. In practice, there is little to no overhead, and a compromise between trip rate and energy reach must be reached.

In order to better understand the nature and frequency of these faults, we developed a new data acquisition system (DAQ)—by taking advantage of the digital LLRF system of the C100s—that simultaneously records waveforms of 17 different rf signals for each of eight cavities in a cryomodule. This process is triggered when the LLRF system for any cavity in the cryomodule detects a fault condition. Typically, these waveforms are recorded at 5 kHz as compared to the conventional archived signals (up to 10 Hz). The stored waveform data provide sufficient resolution that, off-line, a subject matter expert (SME) can identify which cavity faulted first and the type of fault. However, in order to provide near real-time feedback to
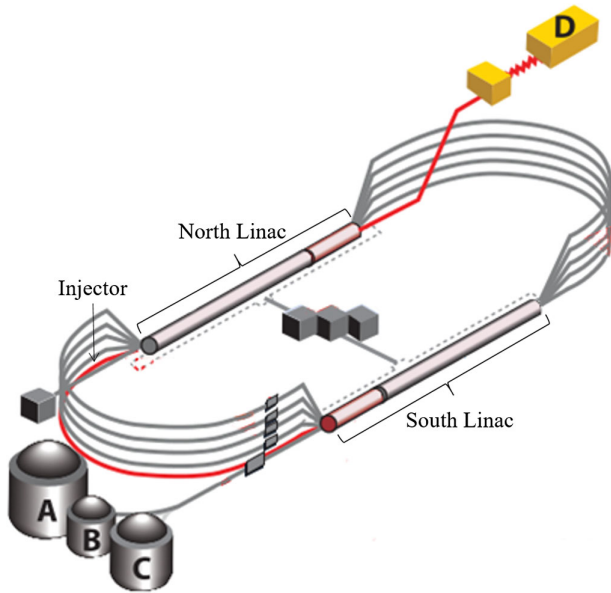
FIG. 1. Schematic of the CEBAF accelerator showing the four experimental halls (A, B, C, and D) and indicating the locations of accelerating sections. The north and south linacs are comprised of 25 cryomodules each. The five most downstream cryomodules of each linac are populated by C100s. The remaining two C100s are located in the injector with one additional one in the north linac.

control room operators, an online system that utilizes machine learning to automatically identify (i) the cavity that faulted first and (ii) the type of fault was developed and a prototype system implemented prior to the fall 2019 physics run. In this paper, we outline the development of the various system components and report on initial results. This represents a type of prognostics problem—a class of problems particularly important for accelerator facilities [3]. Indeed, similar work at other facilities, targeting different machine subsystems and using different techniques, shows the promising potential of machine learning [4,5,6].

In Sec. II, we further motivate this work, and subsequent sections describe the various components that comprise the machine learning system. Specifically, Sec. III describes the data acquisition system, Sec. IV discusses the off-line analysis and the critical step of labeling the data, and Sec. V details how the machine learning models were developed and implemented into a framework for online use. We conclude with results from the winter 2020 physics run and discuss future work.

II. MOTIVATION

The cavities in a C100 cryomodule have strong cavity-to-cavity mechanical coupling. When a cavity is turned off or the gradient is reduced, the change in length due to Lorentz force detuning leads to changes in the length of adjacent cavities. These effects are exacerbated by the mechanical resonances of the structure. The result is that, when one cavity trips off, the remaining seven cavities are affected with several likely to trip from vibrational-induced detuning. This leads to an operational problem with the C100 cavities, namely, answering the question "which cavity tripped first?" Identifying the offending cavity with existing software and hardware is difficult to do automatically, because secondary faults occur with delays on the order of milliseconds. While control room operators have access to the raw waveform data that are captured during a fault, correctly interpreting the signals requires a subject matter expert.

Developing machine learning models to automatically label fault events would immediately reduce 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 the cavity gradient, while another type of fault may indicate a larger problem and require intervention from 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. We have already reaped the benefits of this latter kind of analysis. For example, postmortem analysis of fault data from the fall 2018 physics run indicated three cryomodules in the south linac (2L24, 2L25, and 2L26) were particularly susceptible to microphonic-based faults. This provided the impetus to perform microphonics hardening (installing tuner dampers) on these cryomodules during a down cycle. As a result, in the following physics run, there was a reduction in microphonics-based trip rates, thereby allowing gradients to be increased in those cryomodules as well [7]. In another instance, results of manual fault classification uncovered a firmware bug which caused false trips of a type which represented approximately 40% of the faults. Additionally, results were used to monitor electronic quench fault types to better understand if there was a need to periodically thermally cycle C100 cryomodules to room temperature. Thus, both postfault and postrun analysis by a SME is valuable. At present, only the latter is possible and is a laborious and time-consuming process.

III. DATA ACQUISITION

For each C100 cavity fault, the newly developed data acquisition system synchronously acquires timestamps and saves waveform records of 17 different rf signals from each of the eight cavities in the cryomodule. The data acquisition system is comprised of two primary components, the LLRF and experimental physics and industrial control systems (EPICS), along with a collection of high-level applications. These two components work together to generate and store the data for off-line analysis and
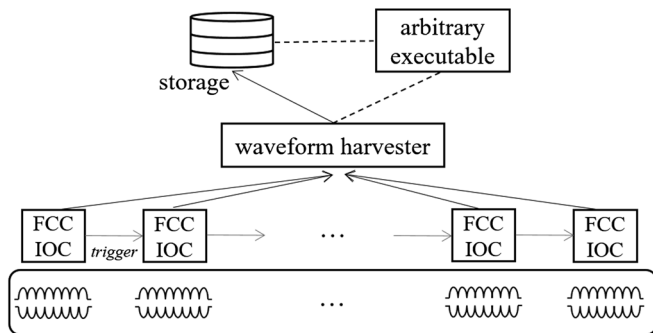
FIG. 2. A schematic showing the data generation and storage systems. A fault at a cavity triggers data collection at that cavity's FCC input or output controller (IOC), and the fault trigger is propagated to neighboring FCC IOCs associated with the cryomodule.
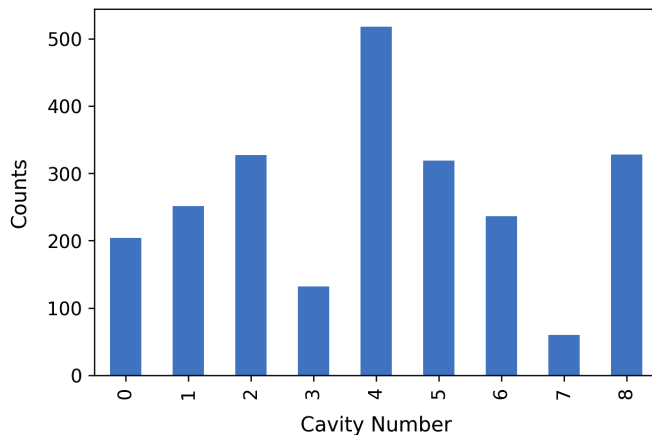

FIG. 3. Histogram showing the distribution of fault events by cavity identification. Numbers 1–8 correspond to unique cavity locations within a cryomodule, and cavity 0 corresponds to instances where all cavities tripped off simultaneously.

inspection. An Altera field programmable gate array (FPGA) is the signal processing engine within the Jefferson Lab 12 GeV rf field control chassis (FCC) [8]. A waveform harvester was developed to capture rf time-series signals after a fault and write them to a file for later analysis. Each of the 17 harvested waveform signals is 8192 points long. The trigger is set such that approximately 94% of the recorded data precede the fault and 6% after. For standard sample rates of 0.20 ms per sample (5 kHz), this yields approximately 1535 ms of data before the fault. These prefault data provide valuable information about the root cause of the trip. After the data have been written to network storage, the harvester calls an external executable. Currently, the executable runs the cavity and fault machine learning models and uploads the information to a centralized database via a waveform-specific web service. Finally, all waveform-related data are kept online indefinitely, backed up to tape daily, and after 90 days are compressed to reduce online storage (see Fig. 2).

IV. DATA ANALYSIS AND LABELING

Since deployment of the data acquisition system in early 2018, an SME has labeled several thousand fault events from C100 cavities. That is, time-series signals from cavities have been assigned corresponding labels indicating the first cavity to trip off and the fault type.

The topic of label reproducibility by the SME will be addressed more fully in Sec. VI; however, it is worthwhile to provide additional background regarding the SME and labeling process, in general. The SME has over 30 years of experience working with SRF cavities and is uniquely qualified to label cavity faults given his work in developing cavity rf systems, diagnostic systems for cavities and cryomodules, and for his SRF operational experience in operational machines and off-line test environments at Jefferson Lab and other facilities, along with his experimental studies investigating cavity faults (i.e., arcing, thermal quenches, and microphonics). He has been the

sole labeler for the project. For the very first labeling exercise (spring 2018), only five fault types were identified. After analyzing data from many hundreds of fault events in subsequent running periods, three additional fault types were identified, yielding the labeled dataset used in this paper with eight fault types. Because we recognize additional fault classes may be required as more data are collected and analyzed, future labeling efforts will include an "other" category for events which do not naturally match preexisting fault types.

For the results in this paper, we consider a dataset comprised of faults recorded during CEBAF beam operations between January 18, 2019, and March 9, 2020. The dataset contains a total of 2375 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. Each signal contains 1638 ms of data sampled at 5 kHz. Details of each signal are listed in the Appendix A. Figure 3 shows the distribution of faults by cavity, where the labels (1,2,3,4,5,6,7,8) correspond to unique cavity locations within a cryomodule and (0) corresponds to instances where all cavities fault, or were turned off, simultaneously. Figure 4 shows the distribution of faults by type.

Possible fault mechanisms are Microphonics, Quench_100 ms, Controls_Fault, E_Quench, Quench_3 ms, Single_Cav_Turn_Off, Heat_Riser_Choke, and Multi_Cav_Turn_Off. Readers interested in a detailed analysis of the time-domain signals and the criteria used for labeling each fault class are directed to Ref. [7]. Below, we provide a brief description of our current best understanding of the mechanism underlying each fault type.

A microphonics fault is cause by detuning of the cavities, generally due to the vibrational modes of the cavity string. Because of the geometry of the structure, the center of the string has a stiffer constraint, and cavities 4 and 5 generally have higher frequency shifts and often will trip first [9].
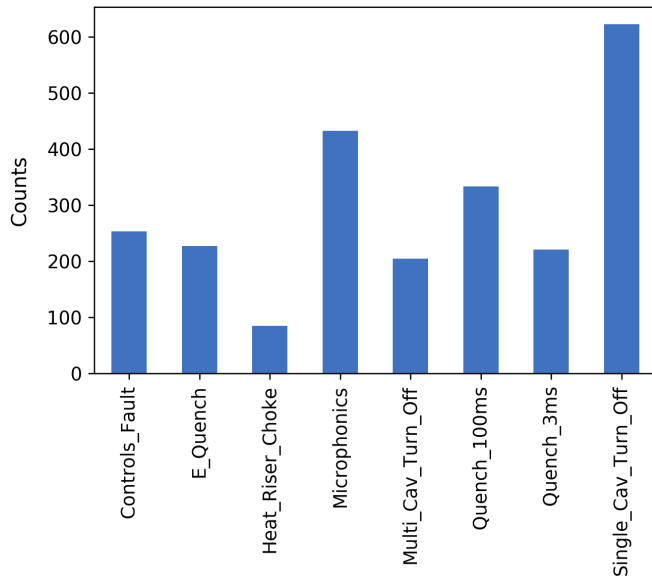
FIG. 4.　Histogram showing the distribution of fault events by type. The data represent a total of 2375 unique C100 cavity fault events.

The 100 ms quench fault ("Quench_100 ms") is assumed to be a thermal quench that is initiated in the end group which is conduction cooled. Detailed analysis done as part of the design process indicated that the propagation time was 100 ms [10]. In these events, the propagation of the quench is characterized by a single cavity having a substantial change in cavity frequency as indicated by a change in the relative phase between the forward rf power and the cavity field (known as the detune phase). A 3 ms quench fault ("Quench_3 ms") is similar in nature albeit with a faster shift in detune phase or an uncontrolled 1% reduction of cavity gradient with timescales on the order of 1 ms. Both the 100 and 3 ms quench faults may exhibit variation in their characteristic timescales; nevertheless, it is not so large that one fault could be mistaken for the other. An electronic quench fault ("E_Quench") refers to an event that involves a gas discharge in the cavity [11]. In these events, the accepted theory is that electrons are stripped off of gas molecules and accelerated by cavity fields. The initiating mechanism has been, and continues to be, an area of investigation [9]. In such events, the stored energy in the cavity, as indicated by the measured gradient, is depleted in times on the order of 10 μs, which is much faster than is possible with a thermal quench. More than 95% of the electronic quenches occur in cavities 1 or 8. Unlike other fault types, an interlock fault ("Single_Cav_Turn_Off") is not characterized by a perturbation in the cavity rf signals prior to turning off. Therefore, it is labeled through a process of eliminating other fault types. These faults can, however, be correlated to a machine interlock turning off a single cavity (e.g., vacuum faults in the rf waveguides, arc test faults, rf window temperature interlocks, and klystron interlocks). Initially, classifying these events required

looking at supplemental data in order to corroborate the suspected mechanism and, therefore, fault type. However, once a fault type was correlated to specific conditions, mining supplemental data was no longer necessary. A control fault has two or three underlying mechanisms and can be attributed to the controls. One example source of this fault is a perturbation in the 70 MHz master oscillator signal which is used in the cavity's field control chassis. Another common example is the several-kilohertz oscillation in the rf drive signal which occurs when the control loop phase offset is not set to the proper value. Heat riser choke faults ("Heat_Riser_Choke") occur when there is a pressure transient inside the cryomodule which can effect multiple cavities either as a microphonicslike frequency fluctuation or as a slow drift [12]. Unlike a microphonics perturbation which builds up over several hundred milliseconds, a heat riser choke typically causes cavity frequency shifts that turn on in a stepwise manner. A multicavity turn-off fault ("Multi_Cav_Turn_Off") occurs when all eight of the cavities are turned off simultaneously with no prior major variation in the rf signals. Typical drivers for such events are when the operators turn the zone off, when there is a beam line vacuum fault, or when there is a fault in the high-power amplifier, which is an eight-klystron assembly.

To better understand the labeling process, consider an example of a microphonics-induced fault. Microphonics describes time domain changes in the frequency of an SRF cavity, generally when the cavity has been perturbed externally. For cavities that have a large loaded Q, changes in length of a few tens of nanometers cause sufficiently large frequency shifts that cause problems. Typically, cavities will vibrate at the modal resonance frequencies of the mechanical structure. For the C100 cryomodules, the modes are the 9.5–10.5 Hz full string mode, 20–21 Hz half string mode, individual 40–45 Hz mode, and the 80–90 Hz tuner stack mode [13]. The cavity vibrations can also be excited by external sources such as 120 Hz from a vacuum pump which may be attached to the insulating vacuum vessel. Exciting the structure at any of these modal resonances amplifies the excitation and may cause the cavity to trip.

The rf signal that indicates detuning is the relative phase between the forward power and the cavity gradient signal. In CEBAF controls nomenclature, this signal is labeled DETA2. Figure 5 depicts the waveforms for a microphonics-induced trip. As indicated by the DETA2 signal, the structure was vibrating in the 10 Hz full string mode. As the mode built up, the rf drive (GASK) and forward power signals (CRFP) increase to compensate for the perturbation. Eventually, the control loop drive voltage signal for cavity 4 reached its maximum value of 10 and the drive phase got lost. In this instance, the phase was such that it drove the cavity gradient down faster than the natural decay time of the system. Once cavity 4 tripped off, the remaining
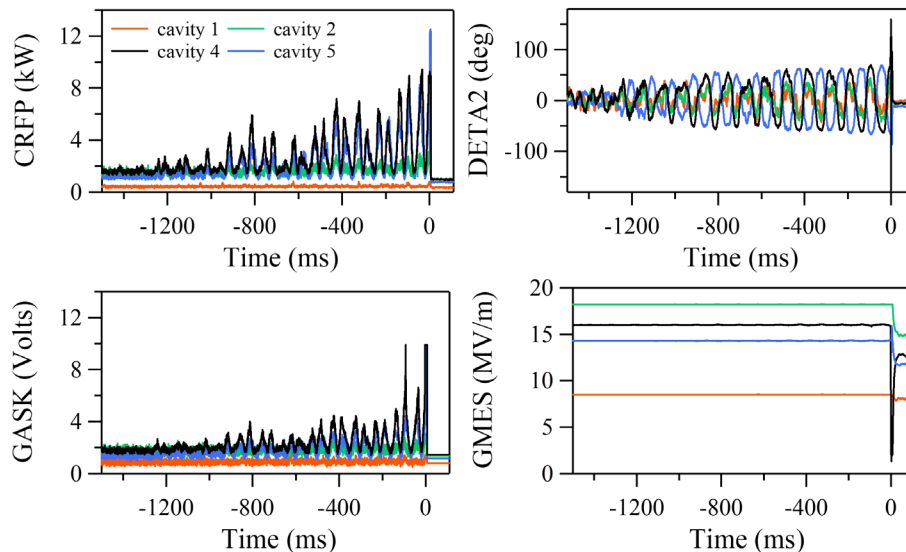
FIG. 5.   Waveforms for a microphonics-induced fault. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). (Only four of the eight cavities are plotted for clarity.)

cavities were switched into self-excited loop mode—a standard protocol with the C100 cavities used to speed up recovery from a trip.

This example illustrates both the domain knowledge required and the laborious nature of labeling fault events by hand. Waveforms associated with the other fault types described in this section are given in Appendix B. While this kind of postmortem analysis is valuable, automating the process for real-time identification of the offending cavity and classification of the fault type would give control room operators actionable information to optimize C100 SRF performance.

V. MACHINE LEARNING MODELS

Machine learning is particularly well suited for applications requiring pattern recognition [14]. The problem we are seeking to solve is one of sequence classification. More specifically, since we are dealing with a series of inputs over time, it is a problem of time-series classification. And because we have labeled examples to train the model, this represents a class of supervised learning. In this section, we briefly outline how the models were developed.

A. Data preprocessing

The values in the raw signal waveforms exhibit large variations (orders of magnitude) between cavities and among signal types within a single cavity. To ensure that the objective function will be well behaved and to encourage faster convergence rates for gradient descent, we apply time-series standardization using the z-score technique on each waveform. The z-score function is shown in Eq. (1):

$$X_{\text{norm}} = \frac{X - \mu}{\sigma},\tag{1}$$

where μ and σ are the mean and the standard deviation of X, respectively.

B. Feature extraction

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. In order to reduce the computational load, features are computed for only four of the 17 recorded rf signals per cavity. These four signals are (GMES, GASK, CRFP, DETA2) and were identified by SMEs as having the highest predictive power for both cavity identification and fault classification. The GMES signal is the calculated accelerating gradient in the cavity, GASK is a representation of a signal in the FPGA proportional to the rf drive voltage, CRFP is the measured forward power, and DETA2 is the relative phase between the rf voltage applied to the cavity and the electric field minus an offset phase (the offset phase is set such that DETA2 is zero when the resonant frequency of the cavity equals the machine reference frequency). Note that we utilize this set of common features for both cavity identification and fault classification models. In this configuration, we perform feature extraction once for each event and maintain two independent classifiers trained for each task using the appropriate class labels. This strategy reduces the computational burden otherwise required to handle independent pipelines for each task separately.

These features can be a set of derived properties from the data that essentially convert the temporal classification to a more amenable static classification task [15]. These intermediate representations could be a set of simple statistics such as mean and variance, skewness, kurtosis, largest peak, and number of zero crossings [16]. More descriptive features such as autoregressive coefficients [17], among others, may be required to obtain a more discriminatory representation of data. An autoregressive (AR) model is a linear predictive model where the next value of a time series is estimated based on previous values of the same time series using linear regression [18]. For instance, estimating $x^{(t)}$ based on $x^{(t-1)}$ is given as

$$x^{(t)} = \varphi^0 + \varphi^1 x^{(t-1)} + \epsilon^t, \tag{2}$$

where φ^0 and φ^1 are regression parameters of the model and ϵ^t is white noise. The model in Eq. (2) is regarded as a first-order autoregressive model [e.g., AR(1)] as the current estimate is based only on the previous value of the time series. The accuracy of the estimates may change based on the amount of historical data we use for the estimation with higher-order AR modeling. A generalization of Eq. (2) for an AR model with order k is given by

$$x^{(t)} = \varphi^0 + \sum_{i=1}^{k} \varphi^i x^{(t-i)} + \epsilon^t. \tag{3}$$

There are many ways to estimate the best parameters φ for a given AR model, such as linear least squares methods. We utilize the ar_model.AutoReg [19] function in the PYTHON-based "statsmodels" library for the AR-based feature extraction scheme. This function estimates the AR parameters φ using the ordinary least squares method [18]. The process of fitting a AR(k) model to a given time series results in a set of parameters, $\Phi = \{\varphi^0, \varphi^1, ..., \varphi^k\}$, that best characterizes the waveform. These parameters become the features for our machine learning models. We select an AR model order of $k = 5$ to fit rf waveforms, as a compromise between feature length and model performance. This results in 192 features (eight cavities per event × four signals per cavity × six features per signal) for each fault event. The same features are used for both cavity identification and fault classification tasks.

C. Model evaluation

Prior to training, we split the 2375 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 [20] library were trained, including k-nearest neighbors, decision tree, support vector, and Gaussian naive Bayes. We also included ensemble
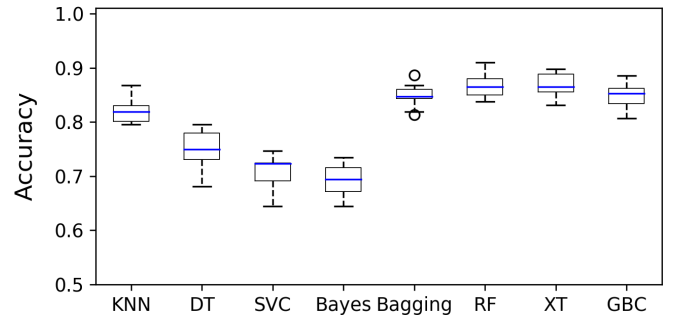

FIG. 6. Boxplots showing accuracy scores from a tenfold cross-validation analysis of several algorithms for cavity identification. The blue line denotes the median, the box spans the interquartile range (IQR), and the upper (lower) whiskers indicate values 1.5 × IQR above (below) the upper (lower) box boundary, with data beyond the whiskers represented as open markers. Ensemble models (four rightmost boxplots) exhibit the best performance.

models such as the bagging classifier, random forest, extra trees, and gradient boosting. Ensemble methods are a machine learning technique that combines several base models in order to produce one optimal predictive model [21]. The basic idea is to combine the predictive power of many "weak learners," and in doing so the error can be dramatically reduced. For each model we use k-fold cross-validation to estimate the performance, ensuring the test data remain untouched. The idea behind k-fold cross-validation is to use training data to generate multiple mini train-test splits and use these splits to tune the model [22]. In standard k-fold cross-validation, the data are partitioned into k subsets, called folds. The algorithm is iteratively trained on $(k - 1)$ folds while using the remaining fold as the test set. Training all of the models as described required less than 5 min of compute time on a laptop (Intel Core i7 processor). Figures 6 and 7 display results of training models for cavity identification and fault classification,
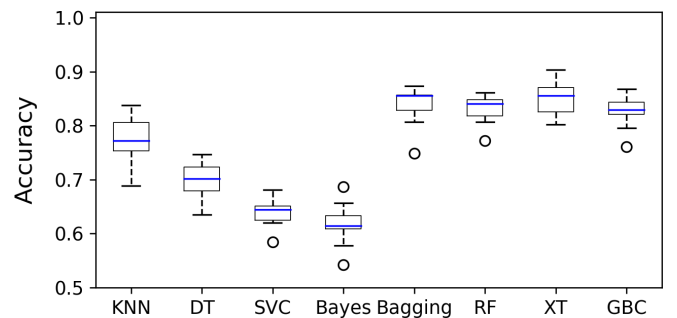

FIG. 7. Boxplots showing accuracy scores from a tenfold cross-validation analysis of several algorithms for cavity fault classification. The blue line denotes the median, the box spans the interquartile range (IQR), and the upper (lower) whiskers indicate values 1.5 × IQR above (below) the upper (lower) box boundary, with data beyond the whiskers represented as open markers. Ensemble models (four rightmost boxplots) exhibit the best performance.

respectively. The accuracy plotted on the vertical axis refers to the ability of the models to reproduce labels given in the training set data. Note that the ensemble models generally outperform the others.

D. Model selection and hyperparameter tuning

The random forest classifier was chosen to model both the cavity identification and fault classification because of its good performance (see Figs. 6 and 7) but also for its robustness against overfitting—a concern when dealing with smaller datasets. The next step is to optimize the model through hyperparameter tuning. Hyperparameters are parameters that are not directly learned but rather are passed as arguments to the constructor of the estimator classes. Rather than do a manual search of the parameter space, we invoke GridSearchCV to automate the process of scanning over specified parameter ranges. Specifically, in the case of the random forest, we scan (*n_estimators, criterion*, and *bootstrap*). Once again, we avoid evaluating the model on the test data and take advantage of tenfold cross-validation to estimate model performance, thereby keeping the test set as a truly unseen dataset for final model evaluation. The hyperparameters corresponding to the best accuracy score define the final model and are given in Table I.

E. Performance metrics

There are a multitude of metrics that can be used to assess the performance of classification models. Initially, the tenfold cross-validation score (based on training data) is compared with predictions from the withheld test set. In particular, we want to avoid to overfitting, marked by a cross-validation score being much greater than the accuracy derived from test data. Table II 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 I. Results of hyperparameter tuning for each random forest model.

	Cavity identification	Fault classification
n_estimators	300	800
Criterion	gini	gini
Bootstrap	False	False

TABLE II. Cross-validation and accuracy scores for each random forest model. The accuracy scores were generated by applying the model on the withheld test dataset.

	Cavity identification	Fault classification
Tenfold cross-validation (%)	87.97 ± 1.81	85.52 ± 3.65
Accuracy (test data) (%)	87.94	87.66

F. Accessibility

Making data publicly available is necessary to drive innovation and progress. In an effort to make data findable, accessible, interoperable, and reusable [23], the feature matrix (2375 instances $\times$ 192 features per instance) and associated cavity identification and fault classification labels used in this publication are publicly available [24]. Furthermore, to make the data more accessible to nondomain scientists, we follow the guidance put forward by Ref. [25] and provide an associated datasheet as documentation. Likewise, the saved random forest models are also available [26]. Analogous to the concept of datasheets for datasets, we follow the recommendations of Ref. [27] and provide a "model card" for each model, which describes characteristics such as the type of model, intended use cases, and performance metrics.

G. System implementation

Following a C100 cavity fault event, software automatically collects rf signals, saves the data to disk, and passes it to the machine learning models. The entire workflow is summarized in Algorithm 1.

Algorithm 1. Machine learning for online cavity and fault recognition.

(i) Initialization
   (a) Import relevant libraries in Python
   (b) Load trained classifier models for cavity identification and fault classification
   (c) Read in the rf signals for current instance
For each cryomodule:
(ii) Preprocessing
   (a) Check for missing waveforms and missing cavities (data not recorded)
   (b) Select the four representative waveforms (GMES, GASK, CRFP, and DETA2) from each cavity and aggregate
   (c) Standardize each waveform using "*z*-score" method in Eq. (1)
(iii) Feature extraction
   (a) Fit an AR(5) [Eq. (3)] model to each of the 32 waveforms
   (b) Extract the parameters $\Phi = \{\varphi^0, \varphi^1, ..., \varphi^5\}$ from each AR model and aggregate to form a $6 \times 32 = 192$ feature vector
(iv) Inference
   (a) Apply trained cavity identification classifier to feature vector and obtain output
   (b) Apply trained fault classification classifier to feature vector and obtain output
   (c) Aggregate output for each cryomodule
End
(v) Report
(a) Save and display machine learning results to the operators

---

The outputs are stored in a database and presented to operators via a web-based rf waveform viewer. Results can be presented in a table, downloaded as a comma separated file, or plotted graphically (see Fig. 8). Offline, a SME labels the event, and these labels are used to compare the model performance and to add to the training data for additional retraining in the future.
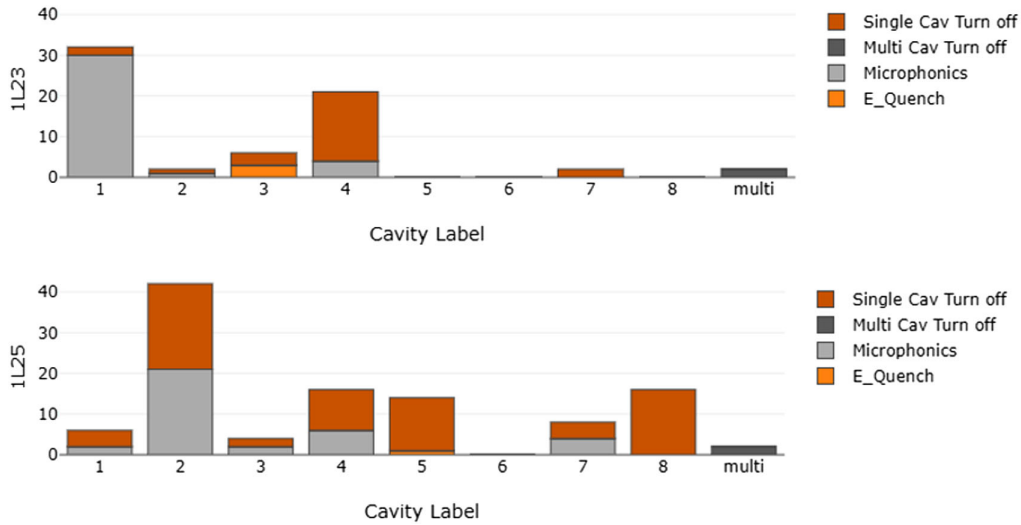
FIG. 8. Partial screenshot showing a graphical summary of machine learning model predictions. Each plot window provides information about the number of faults for an individual cryomodule (vertical axis). Fault types are color coded and displayed for each of the eight cavities (horizontal axis).

VI. RESULTS

Machine learning models were deployed online and used to analyze C100 cavity fault events from March 10 to March 24, 2020—until CEBAF operations ended prematurely due to COVID-19. During that two-week period, 312 fault events were labeled by the machine learning models as well as a SME. The models identified the first cavity to fault and the type of fault with accuracies of 84.9% and 78.2%, respectively. The performance of the cavity identification model is consistent with the performance metrics in Table II; however, the fault classification model underperforms. The models correctly predicted both the cavity and fault for 73.1% of the cases.

While these accuracies provide a general sense of how well the models are performing, a more insightful metric for classifier performance is the confusion matrix. A confusion matrix summarizes the percentage of correct and incorrect predictions by class. It provides an intuitive visualization for which classes are mislabeled (where the model is "confused"), thereby providing insight not only into the errors being made but, more importantly, the types of errors being made. A perfect model would show 0% on all off-diagonal elements and 100% along the diagonal. Figures 9 and 10 show the confusion matrix for the cavity and fault model, respectively. The precision and recall tables, another useful metric, particularly for imbalanced datasets, are given in Appendix C.

Several observations are immediately apparent: (i) The cavity model has particular difficulty identifying cavities 3 and 7. This is not surprising, as these two cavities are the least represented in the training data (see Fig. 3). (ii) The fault model was unable to correctly classify heat riser choke faults. Similar to the previous point, this fault has the smallest number, by far, of training examples (see Fig. 4).

However, there are additional reasons for the model's confusion with regard to this particular fault that will be discussed below. (iii) The fault model is easily confused by controls faults. Discussions with a system expert confirm that, with a few exceptions, it is typically difficult to understand the underlying mechanism of controls faults from the waveform data. This confirms that, when the labeling is difficult, not surprisingly, it is often reflected in poor model performance. (iv) During the two-week model testing in March, there were several clusters of events where the models performed poorly and that warranted closer inspection. These are discussed briefly below. (v) During the evening of March 16, eight heat riser choke faults were identified in
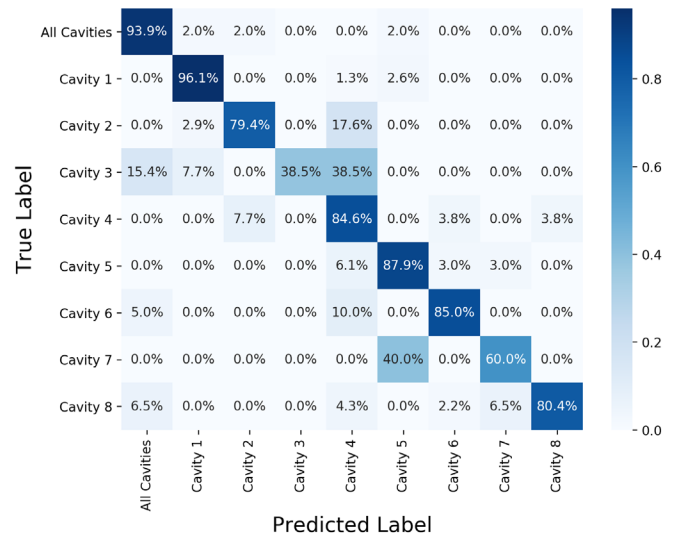

FIG. 9. Confusion matrix showing performance of the deployed cavity identification model on 312 fault events compared to the labels provided by a subject matter expert.
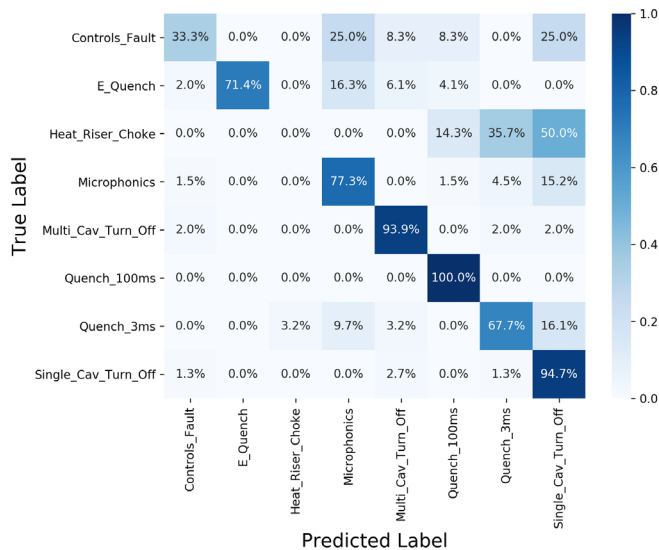
FIG. 10. Confusion matrix showing performance of the deployed fault classification model on 312 fault events compared to the labels provided by a subject matter expert.

cavity 2 in cryomodule 1L23, which the fault model failed to correctly label. These are relatively rare fault types, and having a succession of them over a period of just a few hours represents anomalous behavior. On closer examination of the raw waveforms, the multiple faults are a product of operators not allowing the rf system sufficient time to settle from the previous fault before trying to recover the cavity. Furthermore, the cavities that exhibited the characteristics of a heat riser choke (cavities 2 or 3) were not the cause of excess heat. Rather, it was a thermal quench of cavity 1 which occurred several minutes earlier. The poor model performance may be attributed to the fact that the characteristics of the heat riser fault while recovering the cavity (as opposed to the event occurring during normal beam operation) were not well represented by the already limited training data for this fault type. (vi) On March 18 at 8:02 a.m., all of the C100 cryomodules tripped off simultaneously. This highly unusual event was quickly correlated to a central helium liquefier (CHL) compressor configuration change. In this case, the machine learning cavity model failed to identify the correct cavity for each cryomodule trip. However, this is to be expected, as the CHL event created conditions in the machine that were not represented in the training data and, therefore, the model struggled. (vii) It is worth noting that, if these events are neglected, the cavity model accuracy increases to 86.6%.

Of the 312 events, there were 84 instances where the model disagreed with the labels (either with the cavity identification and/or the fault classification). These events were combined with 84 random events where the model agreed with both labels and submitted to a SME for relabeling. The accuracies of the models assume the labels are ground truth, yet we know there are errors associated with labeling. This was a blind test in that the SME was not given information about the previous labels. The results of the exercise were illuminating. Of the 84 events where the model predictions matched the initial labels, after relabeling by the same SME, of the 168 labels (84 events × two labels per event) only two were labeled differently. On the other hand, of the 84 events where the model predictions differed from the initial labels, the SME relabeled 58 of the 168 labels differently. This result underscores the fact that labeling for this task is difficult. It requires significant experience and intuition about SRF cavities operating with a beam to understand the complex physical mechanisms that give rise to the data collected. And even then, faults can manifest themselves in subtle, but different, ways depending on details of the operating conditions which can vary from run to run and even day to day (i.e., vacuum conditions, gradient distribution, beam loss patterns, and field emission).

When the new labels are compared to the machine learning model predictions, we find that the agreement significantly increases. Table III summarizes the accuracies of each model with each set of labels, original and relabeled. Because only a subset (168 out of 312) of the events were relabeled, we make the assumption that events not relabeled (all of which agree with the models) would remain unchanged. This is reasonable given the 98.8% agreement of the SME when relabeling events where the model agreed the first time.

As a result of the relabeling, the models more closely approach their anticipated performance (Table II). However, more fundamentally, we find that when the models struggle it is because the subject matter expert struggles. This is not surprising. If the SME is less certain about particular kinds of events, it will manifest itself in the training data and, therefore, what the model learns. The model is only as good as the training data it is exposed to.

While the aim is to develop machine learning models with accuracies as high as possible, it should be emphasized that, for this to be a valuable tool for machine operators, the current level of performance is perfectly acceptable. The system provides near real-time information about C100 cavity faults that was previously unavailable. Operators and system experts alike will utilize the information presented to look for trends over time that would necessitate a change in a cavity or cavities. In other words, the effectiveness of the system does not depend on the ability to accurately classify a single, standalone event.

When the models report consistent trips from the same cavity in a cryomodule and of the same fault type, this suggests that intervention by an operator or system expert is

TABLE III. Model accuracies as computed for the original set of labels and after a subset of events—including all that disagreed with the models—had been relabeled.

	Original	Relabeled
Cavity identification	84.9%	88.1%
Fault classification	78.2%	82.4%
Both	73.1%	78.2%

TABLE IV. Problematic cavities as identified by the machine learning models during a two-week test run in March 2020. The cavity identities, number of times the cavity tripped, fault types, and suggested interventions are given.

	Problematic cavities			
Cryomodule	1L23	1L24	1L26	2L26
Cavity identification	1	6	8	4
Number of trips	18	15	23	5
Fault classification	Quench_3 ms	Quench_100 ms	E_Quench	Microphonics
Intervention	Lower gradient	Lower gradient	Lower gradient	Tune cavity

appropriate. Machine learning model predictions from the two-week physics run clearly point to several cavities that required attention (for this initial test, machine learning predictions were not presented to operators). The cavity identities, fault types, and suggested interventions are summarized in Table IV. Together, these four cavities accounted for 19.6% of the faults during the physics run. Applying the appropriate intervention would have had a measurable effect on beam availability to users.

VII. FUTURE WORK

Initial experience with a prototype machine learning system has provided insights into opportunities to further advance this work. These include: (i) *Continue to collect and label data.*—In particular, retraining the models with additional data for cavities and fault types that are not currently well represented is expected to increase accuracy. (ii) *Use unsupervised machine learning techniques to identify fault types.*—At present, a SME determines the fault classes. Therefore, when presented with an unfamiliar pattern in the data, it raises the question of whether it is a variant of a known fault type or if it represents a completely new fault type. Invoking clustering algorithms [28] may provide a data-driven method for identifying the number of fault types represented in the data and potentially reveal new fault types. (iii) *Replace current machine learning models with their deep learning counterparts.*—Deep learning is a subfield of machine learning which is based on learning successive layers of increasingly meaningful representations of the data. The primary advantage of methods based on learning data representations is that it avoids the computationally costly feature extraction step, thereby bypassing a bottleneck in our current workflow. Rather than explicitly compute features, the recorded data serve as the input, and the model learns the necessary features for inference. (iii) Initial studies suggest machine learning can extract information in the signals preceding the fault for prediction, such that it may be possible to design future SRF cavities and their associated control systems with the ability to react quickly enough to avoid the fault altogether [29]. A natural extension of this work is to develop a data pipeline, workflow, and models to stream continuous C100 rf data and provide real-time predictions for impending faults. This represents a critical step toward developing systems that could anticipate and then apply corrective actions to avoid the fault.

VIII. SUMMARY

We have described an implementation of machine learning models at CEBAF to automate the task of classifying C100 SRF cavity faults. The success of these models is a direct consequence of collecting the appropriate data. The importance of having the right data, at the right time, cannot be overemphasized. This would not have been possible without the capabilities of a digital LLRF system, together with a specially developed data acquisition system. Significant investment was necessary to get a functional DAQ, requiring the coordinated effort of multiple groups at the lab. Our experience suggests that, for the potential of machine learning to be fully realized at accelerator facilities, revisiting how and when data are recorded may be required [e.g., modifications to underlying control software to gather data in a way that makes it amenable to machine learning (ML) [30]].

In addition to collecting information-rich data, an equally important contribution was the process of analyzing and manually labeling the data. Although time consuming, the labeling task is an effective way to identify trends and to strategically deploy mitigations to problematic cryomodules and to implement design changes in interlocks. By leveraging machine learning to automate this process, we now have, for the first time, that information available in near real-time.

The next physics run will provide an opportunity to translate model predictions to actionable information for machine operators. For instance, establish a list of guidelines that connects model predictions with operator actions (e.g., "if fault type X happens Y times in the next Z minutes, reduce cavity gradient by N MV/m"). This represents the natural next step for this work; however, initial efforts were hampered by the premature shutdown of CEBAF. In principle, this could be automated in software; however, it is important in this nascent stage of development for operators to gain trust in the machine learning system and allow a human operator to take the decisive action. To this end, we prefer the phrase "augmented intelligence" (rather than "artificial intelligence") to describe the ML models, which serves to emphasize that they provide information on timescales that were previously not possible, in order to guide human decision making.

ACKNOWLEDGMENTS

TABLE V.   List of the 17 rf signals recorded during a cavity trip along with a brief description of each.

Signal name	Description
IMES	Probe in phase voltage magnitude [analog-to-digital converter (ADC) counts]
QMES	Probe quadrature voltage magnitude (ADC counts)
GMES	Measured gradient (MV/m)
PMES	Measured phase (degrees)
IASK	Drive signal in phase voltage magnitude (ADC counts)
QASK	Drive voltage quadrature voltage (ADC counts)
GASK	$sqrt(IASK^2 + QASK^2) \times (10/ADC\,maximum)$
PASK	$tan^{-1}(IASK/QASK)$
CRFP	Forward power (kW)
CRFPP	Forward power measured phase (degrees)
CRRP	Reflected power (kW)
CRRPP	Reflected power phase (degrees)
GLDE	Gradient error (MV/m)
PLDE	Phase error (degrees)
DETA2	Detune angle (CRFPP−PMES−TDOFF)
CFQE2	Cavity frequency error (degrees)
DFQES	Discriminator frequency error

APPENDIX A: RF SIGNALS AND DESCRIPTION

Table V gives a list of the 17 rf signals recorded by the waveform harvester during each C100 cavity trip, along with a brief description. The four signals used for feature extraction and subsequent training of the machine learning models are (GMES, GASK, CRFP, DETA2).

APPENDIX B: WAVEFORMS OF CAVITY FAULT TYPES

Here we include waveforms of the other seven fault types to complement Fig. 5 (microphonics-induced trip). They are: controls fault (Fig. 11), electronic quench (Fig. 12), heat riser choke (Fig. 13), 3 ms quench (Fig. 14), 100 ms quench (Fig. 15), single cavity turn-off (Fig. 16), and multi-cavity turn-off (Fig. 17). Refer to Section IV for a description of the underlying mechanism of each. The black trace in each plot corresponds to the cavity that faulted first, with the exception of the multi-cavity turn off (Fig. 17) wherein all cavities trip simultaneously. The fault
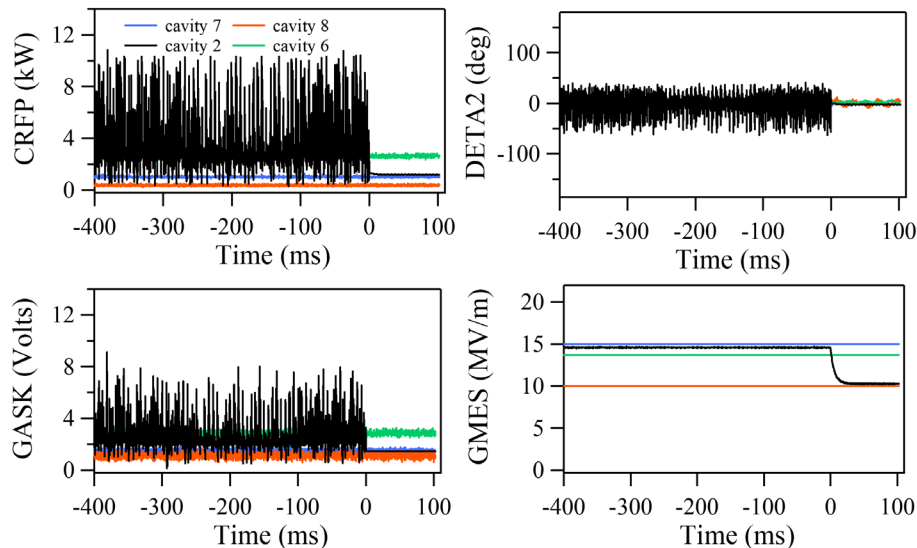

FIG. 11.   Waveforms for a controls fault. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). Only four of the eight cavities are plotted for clarity. Note that the scale of the horizontal axis has been modified to begin 400 ms before the fault event.
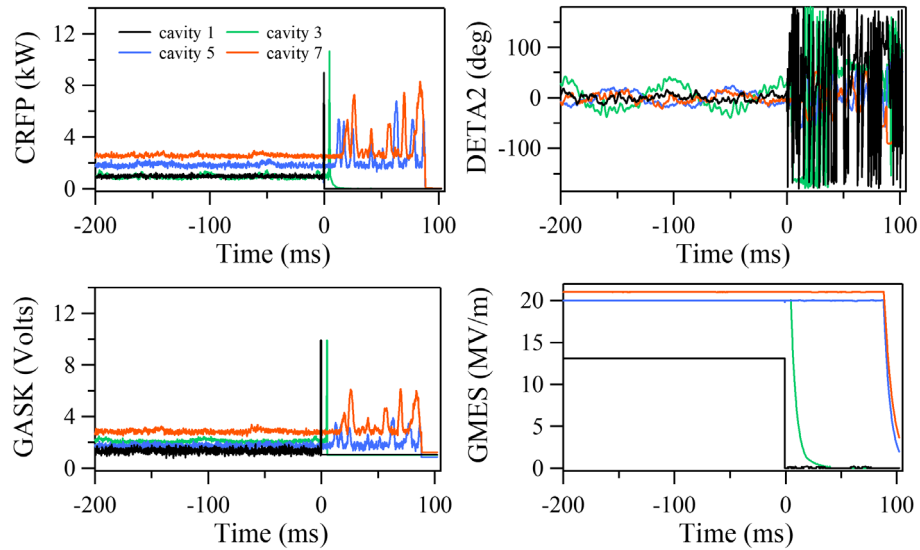
FIG. 12.   Waveforms for an electronic quench. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). Only four of the eight cavities are plotted for clarity. Note that the scale of the horizontal axis has been modified to begin 200 ms before the fault event.
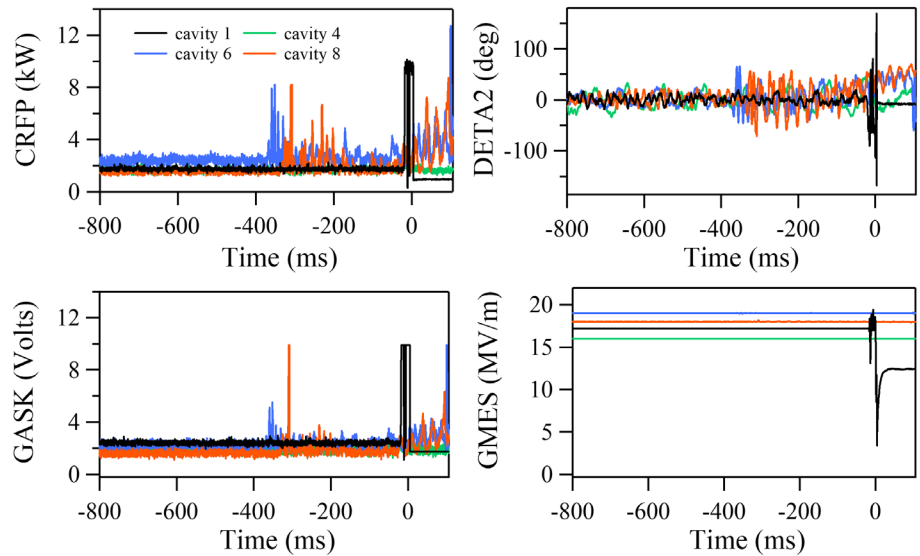

FIG. 13.   Waveforms for a heat riser choke fault. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). Only four of the eight cavities are plotted for clarity.
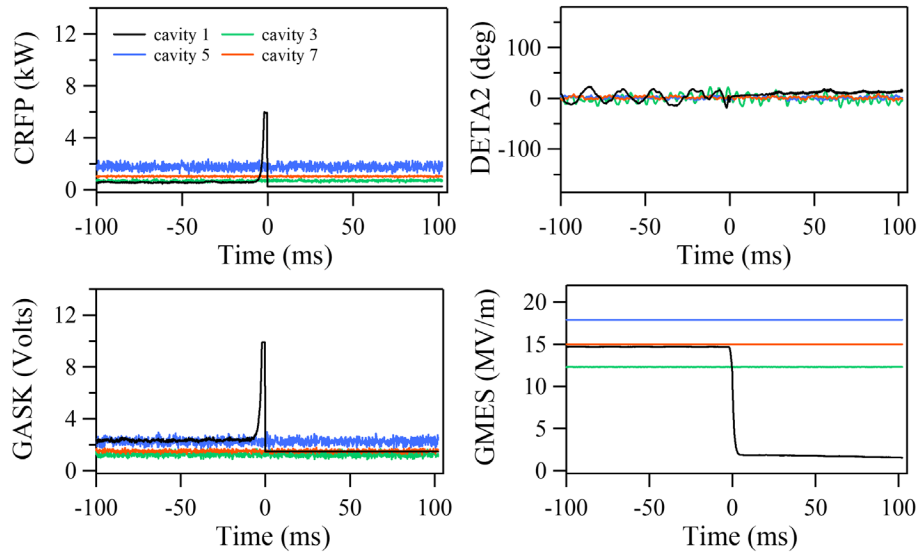
FIG. 14.   Waveforms for a 3 ms quench. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). Only four of the eight cavities are plotted for clarity. Note that the scale of the horizontal axis has been modified to begin 100 ms before the fault event.
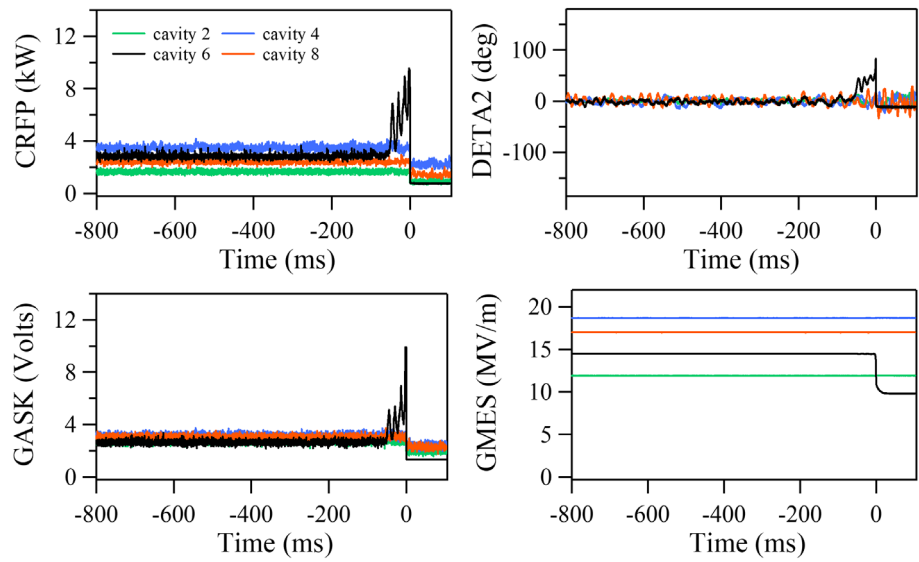

FIG. 15.   Waveforms for a 100 ms quench. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). Only four of the eight cavities are plotted for clarity.
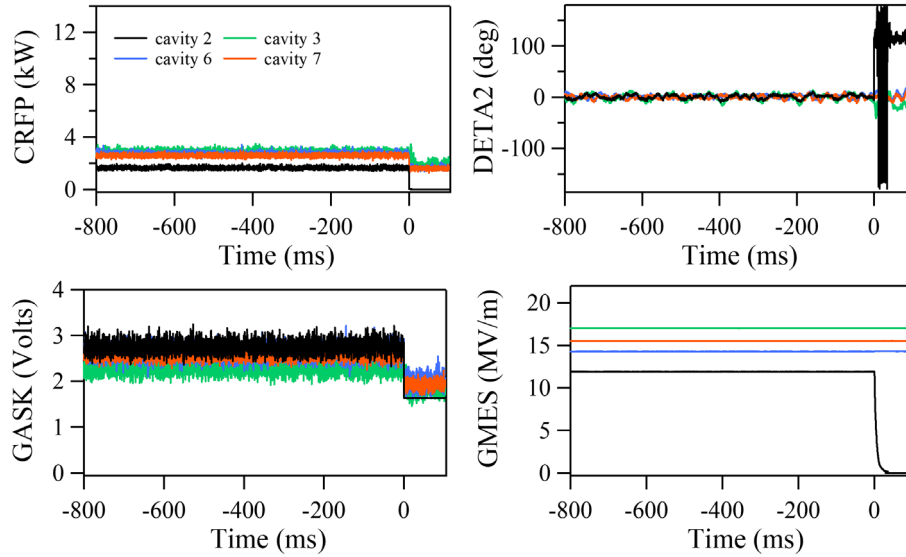
FIG. 16. Waveforms for a single-cavity turn-off. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). Only four of the eight cavities are plotted for clarity. Note that the vertical scale of the GASK signal has been modified.
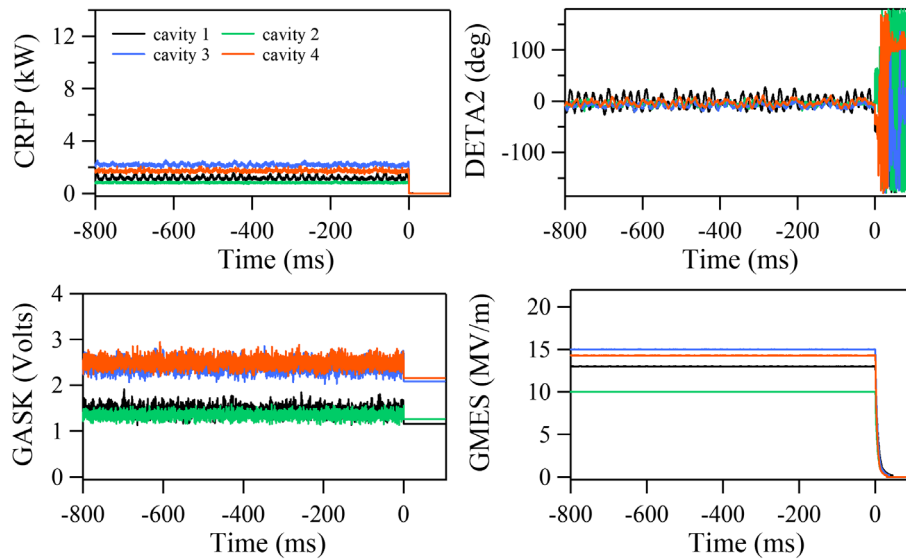

FIG. 17. Waveforms for a multicavity turn-off. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). Only four of the eight cavities are plotted for clarity. Note that the vertical scale of the GASK signal has been modified.

event occurs at $t = 0$ ms and unless noted otherwise, the horizontal scale has been cropped to start 800 ms before, and extend 100 ms after, the fault.

APPENDIX C: PRECISION AND RECALL TABLES

Tables VI and VII display the precision and recall tables for the cavity identification and fault classification models, respectively. These are useful metrics, particularly for imbalanced datasets, and complement the confusion matrices in Figs. 9 and 10. Precision gives the fraction of positive events that were predicted correctly and is given by the ratio of true positives to the sum of true positives and false positives. Recall is the fraction of positive instances that are correctly detected and is given by the ratio of true positives to the sum of true positives and false negatives.

TABLE VI. Precision and recall table for the deployed cavity identification model on 312 fault events from March 10–24, 2020.

Cavity identification	Examples	Precision	Recall
0	49	0.88	0.94
1	76	0.96	0.96
2	34	0.90	0.79
3	13	1.00	0.38
4	26	0.55	0.85
5	33	0.76	0.88
6	20	0.85	0.85
7	15	0.69	0.60
8	46	0.97	0.80

TABLE VII. Precision and recall table for the deployed fault identification model on 312 fault events from March 10–24, 2020. Note that during this time the model did not predict any heat riser choke faults, which is why the precision and recall are both zero.

Fault type	Examples	Precision	Recall
Controls	12	0.50	0.33
Electronic quench	49	1.00	0.71
Heat riser choke	14	0.00	0.00
Microphonics	66	0.78	0.77
Multicavity	49	0.87	0.94
100 ms quench	16	0.73	1.00
3 ms quench	31	0.68	0.68
Single cavity	75	0.73	0.95

[1] C. Reece, Continuous wave superconducting radio frequency electron linac for nuclear physics research, Phys. Rev. Accel. Beams **19**, 124801 (2016).

[2] P. Adderly *et al.*, 12 GeV CEBAF, Phys. Rev. Accel. Beams (to be published).

[3] A. Edelen *et al.*, Opportunities in machine learning for particle accelerators, arXiv:1811.03172.

[4] M. Rescic, R. Seviour, and W. Blokland, Predicting particle accelerator failures using binary classifiers, Nucl. Instrum. Methods Phys. Res., Sect. A **955**, 163240 (2020).

[5] M. Wielgosz, A. Skoczeń, and M. Mertik, Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets, Nucl. Instrum. Methods Phys. Res., Sect. A **867**, 40 (2017).

[6] A. Nawaz, S. Pfeiffer, G. Lichtenberg, and H. Schlarb, Self-organized critical control for the European XFEL using black box parameter identification for the quench detection system, in *Proceedings of the 3rd Conference on Control and Fault-Tolerant Systems (SysTol), Barcelona* (IEEE, New York, 2016), pp. 196–201.

[7] T. Powers and A. S. Shabalina, CEBAF C100 fault classification based on time domain rf signals, in Proceedings of the 19th International Conference on RF Super-

conductivity (SRF'19), Dresden, Germany, 2019, paper WETEB3, pp. 763–769, https://doi.org/10.18429/JACoW-SRF2019-WETEB3.

[8] C. Hovater *et al.*, Status of the CEBAF energy upgrade rf control system, in *Proceedings of the 25th Linear Accelerator Conference (LINAC'10) Tsukuba, Japan* (JACoW, Geneva, 2010), paper MOP095, pp. 280–282.

[9] T. Powers and K. Davis, Transient ponderomotive effects in superconducting cavities, in *Proceedings of the 12th International Workshop on RF Superconductivity (SRF'05), Ithaca, NY* (2005), pp. 619–621.

[10] E. Daly, REN FPC end group transient thermal analysis, Jefferson Lab informal internal presentation, 2005.

[11] T. Powers *et al.*, Investigations of arcing phenomena in the region near CEBAF rf windows at 2 K, in *Proceedings of the 6th International Workshop on RF Superconductivity (SRF'93), Newport News, VA* (1993), pp. 1–18.

[12] S. Van Sciver, He II heat transfer, in *Helium Cryogenics* (Plenum, New York, 1986), pp. 141–155.

[13] K. Davis, J. Matalevich, T. Powers, and M. Wiseman, Vibration response testing of the CEBAF 12 GeV upgrade cryomodules, in *Proceedings of the 26th Linac Conference (LINAC'12), Tel-Aviv, Israel* (2012), pp. 240–242.

[14] C. Bishop, *Pattern Recognition and Machine Learning* (Springer-Verlag, Berlin, 2006).

[15] L. Wang, X. Wang, C. Leckie, and K. Ramamohanarao, Characteristic-based descriptors for motion sequence recognition, in *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining* (Springer, New York, 2008), pp. 369–380.

[16] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, Feature-based classification of time-series data, Int. J. Comput. Res. **10**, 49 (2001).

[17] X. Wang, A. Wirth, and L. Wang, Structure-based statistical features and multivariate time series clustering, in *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM 2007)* (IEEE, New York, 2007), pp. 351–360.

[18] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control* (Wiley, New York, 2015).

[19] statsmodels.tsa.ar_model. AutoReg, https://www.statsmodels.org/stable/generated/statsmodels.tsa.ar_model.AutoReg.html. AutoReg.html (accessed April 21, 2020).

[20] scikit-learn, https://scikit-learn.org/stable (accessed May 13, 2020).

[21] L. Rokach, Ensemble-based classifiers, Artificial Intelligence Review **33**, 1 (2010).

[22] Cross-validation: Evaluating estimator performance, https://scikit-learn.org/stable/modules/cross_validation.html (accessed October 30, 2020).

[23] M. D. Wilkinson *et al.*, The FAIR guiding principles for scientific data management and stewardship, Sci. Data **3**, 160018 (2016).

[24] https://github.com/JeffersonLab/AI_SRF_operations/tree/master/datasets/C100-2020-04-30 10.14462/MLFault-Classifier/1616675.

[25] T. Gebru *et al.*, Datasheets for datasets, arXiv:1803.09010v7.

[26] https://github.com/JeffersonLab/AI_SRF_operations/tree/master/models/C100-2020-04-30.

[27] M. Mitchell *et al.*, Model cards for model reporting, arXiv:1810.03993v2.

[28] D. Xu and Y. Tian, A comprehensive survey of clustering algorithms, Ann. Data. Sci. **2**, 165 (2015).

[29] C. Tennant, On the possibility of using machine learning to predict C100 cavity faults, Jefferson Lab Technical Note No. 20–011, 2020.

[30] G. White *et al.*, The EPICS software framework moves from controls to physics, in *Proceedings of the 10th International Particle Accelerator Conference (IPAC19), Melbourne, Australia* (JACoW, Geneva, 2019), pp. 1216–1218.