

# SUPERCONDUCTING RADIO-FREQUENCY CAVITY FAULT CLASSIFICATION USING MACHINE LEARNING AT JEFFERSON LABORATORY\*

C. Tennant<sup>†</sup>, A. Carpenter, T. Powers, L. Vidyaratne, Jefferson Laboratory, Newport News, USA  
K. Iftexharuddin, Md. Monibor Rahman, Old Dominion University, Norfolk, USA  
A. Shabalina, STFC Daresbury Laboratory, Warrington, UK

## Abstract

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. Of the 418 SRF cavities in CEBAF, 96 are designed with a digital low-level RF system configured such that a cavity fault triggers recordings of RF signals for each of the eight cavities in the cryomodule. Subject matter experts 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 post-mortem – 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. We also discuss efforts for further insights into fault types through unsupervised learning techniques, and present preliminary work on cavity and fault prediction using data collected prior to a failure event.

## INTRODUCTION

The Continuous Electron Beam Accelerator Facility (CEBAF) at Jefferson Lab is a continuous-wave recirculating linac [1]. In September 2017 CEBAF completed an energy upgrade to extend its energy reach from 6 GeV to 12 GeV. To meet this milestone, 88 newly designed cavities (representing 11 cryomodules) were installed. Each cryomodule (comprised of eight 7-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. In order to better understand the nature and frequency of C100 cavity faults, a data acquisition system was implemented to record data for off-line analysis. For every C100 cavity trip, the system automatically records RF signals from each of the 8 cavities in the cryomodule (see Fig. 1). The recorded time-series data allows subject matter experts to analyze the data off-line. Specifically, each event receives two labels; (1) the cavity which went unstable first and (2) the type of fault. This kind of labelled data permits the use of supervised, machine learning (ML) models for classification. Near real-

time identification of the offending cavity and classification of the fault type would give operations valuable feedback for corrective action planning [2]. Additionally, post-run analysis using the aggregate statistics and breaking down faults according to cryomodule and cavity provides data-driven guidance for maintenance and/or upgrades.

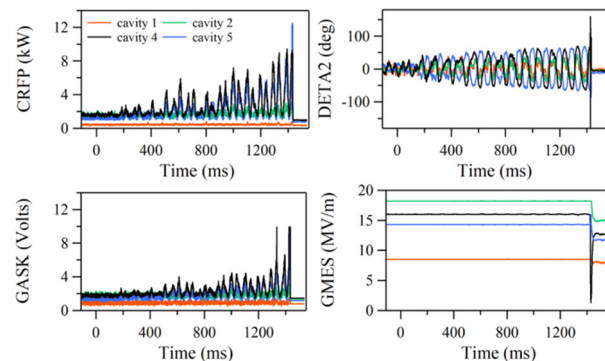


Figure 1: Example waveforms for a microphonics-induced fault. The plots display the forward power (top left), detune phase angle (top right), a digital signal proportional to the drive voltage (bottom left), and the measured gradient (bottom right). (Only four of the eight cavities are plotted for clarity.)

## CAVITY FAULT CLASSIFICATION

Machine learning is particularly well suited for applications requiring pattern recognition. The problem we are seeking to solve is time-series classification. And because we have labelled examples to train the model, this represents a class of supervised learning. In this section we briefly outline how the models were developed.

### *Supervised Learning: Machine Learning*

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 4 of the 17 recorded RF signals per cavity. These intermediate representations could be simple statistics such as mean and variance, skewness, kurtosis, largest peak and number of zero crossing. More descriptive features such as autoregressive coefficients, among others, may be required to obtain a more discriminatory representation of data. For our models we computed 6 autoregressive coefficients for

\*Work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Contract No. DE-AC05-06OR23177.

<sup>†</sup>tennant@jlab.org

each of the 4 waveforms for every cavity in the cryomodule. A single fault event is then represented by a vector of length 192 (6 coefficients  $\times$  4 waveforms/cavity  $\times$  8 cavities/cryomodule). Note, that in an effort to make data Findable, Accessible, Interoperable and Reusable (FAIR) [3], the feature matrix (2,375 instances  $\times$  192 features/instance) and associated cavity identification and fault classification labels used in this publication are publicly available [4].

Prior to training we split the labelled 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 [5] library were trained, including ensemble 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 [6]. 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 remains untouched. Results for the fault classification model are shown in Fig. 2. 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 inability of a model to generalize learning to data outside the training set). Table 1 summarizes the results and indicates that overfitting is avoided.

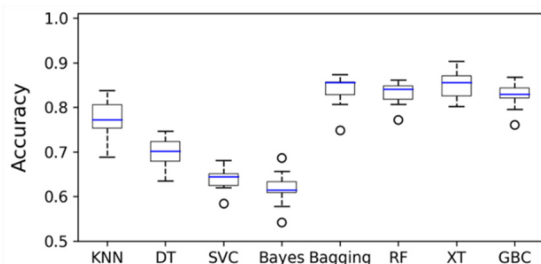


Figure 2: Boxplots showing accuracy scores from a tenfold cross-validation analysis of several algorithms for cavity fault classification. The blue line denotes the median, the box spans the interquartile range (IQR), and the upper (lower) whiskers indicate values  $1.5 \times$  IQR above (below) the upper (lower) box boundary with data beyond the whiskers represented as open markers. Ensemble models (four rightmost boxplots) exhibit the best performance.

Table 1: Cross-validation (CV) 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 CV (%)	88.0 $\pm$ 1.8	85.5 $\pm$ 3.7
Accuracy (%)	88.0	87.7

**Deployed Performance** 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 labelled by the machine learning models as well as a subject matter expert (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 1, however, the fault classification model underperforms. The models correctly predicted both the cavity and fault for 73.1% of the cases.

**Visualization** In order for the ML models to be effective the information they provide must be communicated clearly and concisely. Care was taken to create a visual interface for operators and SMEs that provides both spatial and temporal information about cavity faults. A screenshot is shown in Fig. 3 where the top plot is a timeline of cavity faults color-coded by type and where vertical position denotes cavity number (for a given cryomodule), while the lower plot shows a heatmap of faults versus cavity labels (for a given cryomodule).

### Supervised Learning: Deep Learning

Deep learning is a sub-field 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.

A neural network that contains one or more feedback connections among neurons is known as a recurrent neural network (RNN) and is often used in conjunction with sequences of data. Generic recurrent neural networks face a known problem of vanishing gradient problem over time, similar to the vanishing gradient problem that occurs over depth of a deep network architecture [7]. The long short-term memory (LSTM) unit is developed to address the vanishing error signal, with the introduction of memory gates that control the flow of context over time [8].

One of the model architectures investigated uses 32 time-series signals as inputs (4 time-series/cavity  $\times$  8 cavities) down sampled to 256 time steps. The model contains 3 LSTM layers each with 64 feature dimensionality at the front end for time-series feature learning. The LSTMs are designed with the ability to learn both long-term and short-term temporal features. The back end of the model is a branched architecture with multiple feed-forward neural layers to enable simultaneous learning of both cavity and fault identification tasks. The features learned by LSTMs are shared among the two paths to perform both tasks in a computationally efficient manner. Results with this model yield test accuracies for cavity identification and fault classification of 87.7% and 81.3%, respectively. We note in passing that several convolutional neural network (CNN) architectures were also investigated, some of which yielded comparable performance to the RNN [9].

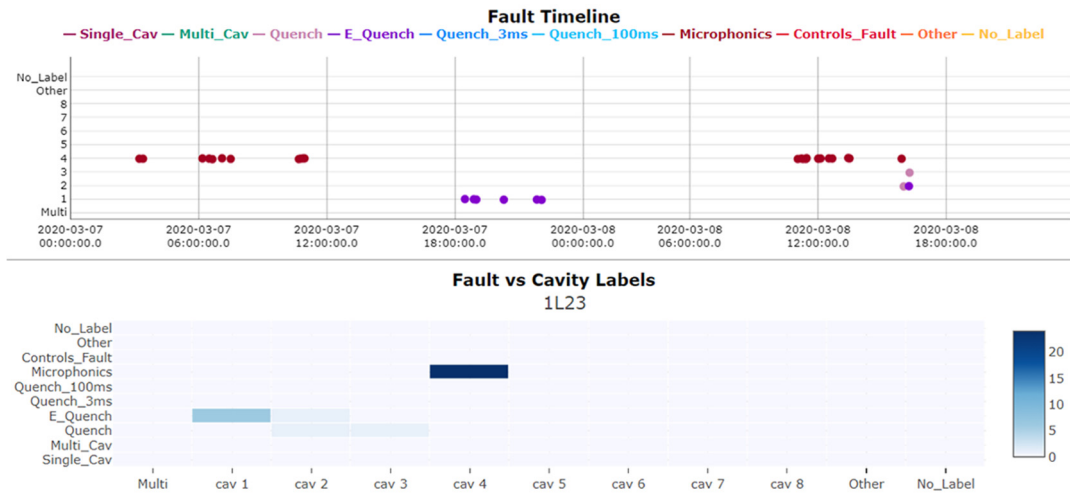


Figure 3: Screenshot of the interactive dashboard created to communicate results of the cavity fault classifier. The top plot is a timeline where markers represent fault events and are color-coded according to fault type and their vertical position denotes the cavity that faulted first. The bottom plot is a heatmap showing the frequency of faults in cavities for a specific cryomodule (1L23).

**Uncertainty Quantification** Before operators take action based on information from a ML model, we need a measure of confidence of the model’s prediction. Monte Carlo dropout is a method that provides uncertainty quantification in deep learning models and can be implemented in a straightforward way [10]. Dropout is a popular regularization method used to avoid overfitting. The idea is to randomly disable some connections in the network for each training example, thereby allowing the model to generalize better. Typically, dropout is only applied during training, whereas in Monte Carlo dropout, it is applied at both training and test time. Consequently, at test time the prediction is no longer deterministic, since random nodes are being disabled. That is, given the same input, the model predicts slightly different values each time. The result of Monte Carlo dropout is to generate random predictions and to interpret them as samples from a probabilistic distribution. Figure 4 displays uncertainty quantification results for a fault classification model. Note that a majority of the confidence values are close to 1 when the model agrees with the ground truth whereas there are a significant fraction of events with confidence below 0.7 when the model disagrees the ground truth.

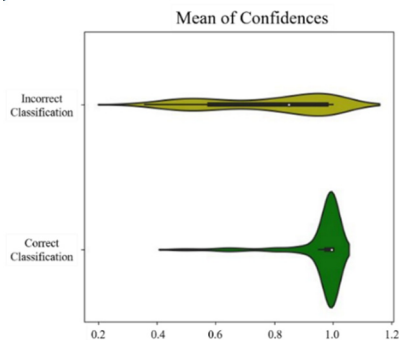


Figure 4: Violin plots comparing the distribution of the mean confidence values when the model agrees with the ground truth labels (bottom) and when the model disagrees (top) when applied to a test set containing 759 samples.

### Unsupervised Learning

Data in supervised learning tasks consists of input-output pairs. The goal of supervised machine learning is to learn the mapping between inputs and outputs such that it can generalize and make inferences on unseen data. The goal of unsupervised learning is to analyze the features (often without regard for the labels) in order to find patterns or underlying structure in the dataset. Within unsupervised learning, there are two common applications; dimensionality reduction and clustering. We focus on the former. The dimensionality of a dataset describes the number of features needed to describe it. Dimensionality reduction provides a means of compressing the features into a lower dimensional space. One primary reason for doing this is to allow for two-dimensional visualizations. For this work we use the *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) algorithm [11]. This kind of analysis has yielded interesting insights into the data. See, for example, Fig. 5 which shows the result of applying *t*-SNE to only fault types labelled by the SME as “single cavity turn offs”. Note the distinct cluster that represents events from one particular cavity in cryomodule 1L23 and is separated from the larger, more uniformly populated cluster. A closer examination of the raw data, in particular the frequency content of the signals, revealed that the 1L23 events exhibited 50 Hz microphonics from an unknown source, whereas microphonic frequencies in the other cryomodules had content at rough 0, 20, 40, and 90 Hz with peak values that are generally 2 to 4 times larger. This may be due to the fact that the 1L23 cavities were a low loss cell shape which is more sensitive to microphonics. Another motivation for using unsupervised learning techniques 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-

Content from this work may be used under the terms of the CC BY 3.0 licence (© 2021). Any distribution of this work must maintain attribution to the author(s), title of the work, publisher, and DOI

driven method for identifying the number of fault types represented in the data, and potentially reveal new fault types.

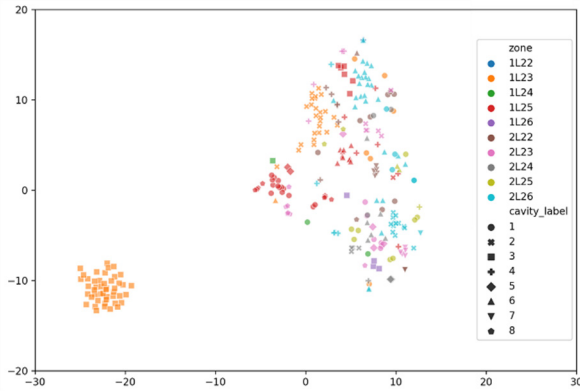


Figure 5: *t*-SNE visualization of all “single cavity turn off” fault types. The events originating in cavity 3 of cryomodule 1L23 comprise a separate cluster. The difference was traced to the frequency response of this cavity due to a different cell shape compared to the other events.

### CAVITY FAULT PREDICTION

Initial studies suggest machine learning can extract information in the signals preceding the fault for prediction [12]. A natural evolution of the work outline in the previous section 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 towards developing systems that could anticipate and then apply corrective actions to avoid the fault.

#### Binary Classifier

As a preliminary study, a dataset was constructed using a 100 ms window of time-series data immediately before a fault (labelled “unstable”) and at 1.5 seconds prior to a fault (labelled “stable”) gathered from over 5,000 saved fault events. The data was used to train a binary classifier. Initial accuracy was very poor (74%) with a large number of false negatives (i.e. the model incorrectly identifies “unstable” data as “stable”). After discussions with a domain expert, it was discovered that several fault types have no precursors – in our data – that a trip will occur. For these faults, ancillary data is oftentimes used to determine the type. After removing instances of faults without precursors, the binary classifier was re-trained and achieved over 92% accuracy. In addition to providing key insights into the behavior of several fault types, the study provided motivation to investigate whether data prior to a failure event could also predict the type of fault.

#### Fault Type Prediction

For this study, a window of a specified length slides over the data starting from the 1.5 second prior to the fault up to the fault itself ( $t = 0$  seconds). At regular intervals a deep learning model is trained using the window of data and the known fault type label. Figure 6 shows the accuracy of

these models plotted as function of the time prior to a microphonics initiated faults. Initial results suggest that time-scales may be favorable in allowing a deep learning model to anticipate several kinds of faults and deploy mitigation measures. There is therefore strong motivation to continue research into this area. While the direct impact to Jefferson Lab, and CEBAF in particular, is constrained by existing cavity and control system designs, this work has the potential for future SRF-driven accelerator facilities.

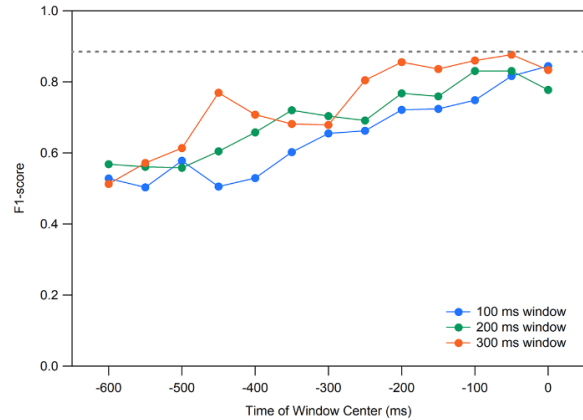


Figure 6: F1-scores [13] of a fault classification model for predicting a microphonics-based event using a window of data as a function of time prior to the fault event ( $t = 0$ ).

### LESSONS LEARNED

The success of all the models discussed in this paper 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. Our experience suggests that for the potential of machine learning to be fully realized at accelerator facilities, revisiting how and when data is recorded may be required. In addition to collecting information-rich data, an equally important contribution is the process of exploring and manually labeling the data. Although time consuming, having a willing and able “data labeler” is critical. We also discovered that models struggle precisely where the subject matter expert struggles. This is not surprising. If a SME is confused about certain kinds of events, it will manifest itself in the training data and consequently, what the model learns.

### SUMMARY

We have described an implementation of machine learning models at CEBAF to automate the task of classifying C100 SRF cavity faults providing, for the first time, that information in near real-time. 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. Recent modifications to the LLRF system will provide access to streaming signals in the near future, providing an opportunity to adapt machine learning models to streaming data and predict faults before they occur (see Fig. 7).

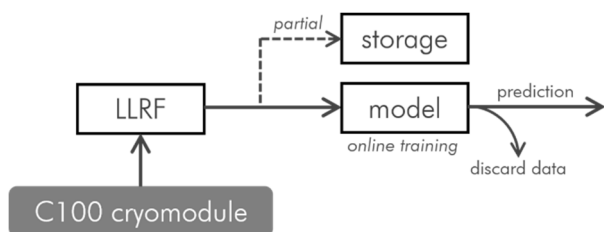


Figure 7: Schematic showing the data flow for a deep learning model with a continuously streaming input. A small portion of the streaming data is sent to storage for off-line analysis.

## REFERENCES

- [1] C. Reece, “Continuous wave superconducting radio frequency electron linac for nuclear physics research”, *Phy. Rev. ST Accel. Beams*, vol. 19, p. 124801, 2016.  
doi:10.1103/PhysRevAccelBeams.19.124801
- [2] C. Tennant *et al.*, “Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory”, *Phy. Rev. ST Accel. Beams*, vol. 23, p. 114601, 2020. doi:10.1103/PhysRevAccelBeams.23.114601
- [3] M. D. Wilkinson *et al.*, “The FAIR guiding principles for scientific data management and stewardship”, *Sci. Data*, vol. 3, p. 160018, 2016. doi:10.1038/sdata.2016.18
- [4] GitHub,  
[https://github.com/JeffersonLab/AI\\_SRF\\_operations/tree/master/datasets/C100-2020-04-30](https://github.com/JeffersonLab/AI_SRF_operations/tree/master/datasets/C100-2020-04-30)
- [5] Scikit-learn, <https://scikit-learn.org/stable>
- [6] L. Rokach, “Ensemble-based classifiers”, *Artificial Intelligence Review*, vol. 33, pp. 1-39, 2010.  
doi:10.1007/s10462-009-9124-7
- [7] Y. Bengio, P. Simard and P. Frasconi, “Learning long-term dependencies with gradient descent is difficult”, in *IEEE Transactions on Neural Networks*, vol. 5, pp. 157-166, 1994.  
doi:10.1109/72.279181
- [8] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, *Neural Computation*, vol. 9, pp. 1735-1780, 1997. doi:10.1162/neco.1997.9.8.1735
- [9] L. Vidyaratne *et al.*, “Deep Learning-Based Superconducting Radio-Frequency Cavity Fault Identification at Jefferson Laboratory”, *Frontiers, Research Topic: Operational Intelligence*, to be published.
- [10] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”, in *Proc. of 33<sup>rd</sup> International Conference on Machine Learning*, New York, USA, Jun. 2016, paper PMLR48, pp. 1050-1059.
- [11] L. van der Maaten and G. Hinton, “Visualizing Data using t-SNE”, *Journal of Machine Learning Research*, vol. 9, pp. 2579-2605, 2008.
- [12] C. Tennant, “Using Unsupervised Learning to Analyze C100 Fault Type Data”, Jefferson Lab, Newport News, USA, Tech. Note 20-020, 2020.
- [13] Scikit Learn,  
[https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\\_score.html](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html)