ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-WEPV021

MACHINE LEARNING FOR RF BREAKDOWN DETECTION AT CLARA

A. E. Pollard*, A. J. Gilfellon[1], D. J. Dunning
ASTeC, Cockcroft Institute, STFC Daresbury Laboratory, Warrington, UK
[1]also at The University of Manchester, Manchester, UK

Abstract

Maximising the accelerating gradient of RF structures is fundamental to improving accelerator facility performance and cost-effectiveness. Structures must be subjected to a conditioning process before operational use, in which the gradient is gradually increased up to the operating value. A limiting effect during this process is breakdown or vacuum arcing, which can cause damage that limits the ultimate operating gradient. Techniques to efficiently condition the cavities while minimising the number of breakdowns are therefore important. In this paper, machine learning techniques are applied to detect breakdown events in RF pulse traces by approaching the problem as anomaly detection, using a variational autoencoder. This process detects deviations from normal operation and classifies them with near perfect accuracy. Offline data from various sources has been used to develop the techniques, which we aim to test at the CLARA facility at Daresbury Laboratory. These techniques could then be applied generally.

INTRODUCTION

There are two main aims with this project. Firstly, we aim to assemble a machine learning (ML) based system that could be used to replace the current mask method of radio frequency (RF) breakdown (BD) detection which is standard in the automated code used in the RF conditioning of accelerating cavities. Secondly, we aim to ensure that the mid-process features of the same mechanism could be used as inputs for an ML algorithm designed to predict whether or not the next RF pulse would lead to a BD.

To this end, we constructed a β convolutional variational autoencoder (βCVAE)[1] with RF conditioning data as inputs. After being trained as an anomaly detector this acted as a live BD detector, in conjunction with a dense neural network (NN), which would act with the capacity to replace the current non-ML based BD detection system. In addition to this, the βCVAE's latent space could act as a viable input for a long short-term memory (LSTM) recurrent neural network (RNN) that could be used to predict BDs, based on the methodology set out by Kates-Harbeck et al.[2] who had success in predicting disruptive instabilities in controlled fusion plasmas.

For this investigation, we used data from the CLARA accelerator (Compact Linear Accelerator for Research and Applications) based at Daresbury Laboratory. CLARA is a dedicated accelerator test facility with the capacity to deliver high quality electron beams for industry and research. In addition to the CLARA data, a larger dataset was provided by the CLIC team at CERN covering a cavity test which took place in CERN's XBOX-2 test stand. The structure tested in this dataset was a T24 high-gradient prototype X-band cavity produced at the Paul Scherrer Institute; further details of this design have been reported previously [3, 4]. The CLARA data was collected as part of the routine RF breakdown detection system.

* amelia.pollard@stfc.ac.uk

RELATED WORK

Solopova et al.'s [5] application of a decision tree model to assign both a fault type and cavity-specific location to a collected breakdown signal at CEBAF represents the first foray into using machine learning to classify RF cavity faults. This work was then continued in Tennant et al.[6] where the authors applied a random forest model to the classification of faults and cavity identity for a larger dataset of breakdown events.

Obermair et al. [7] took the first step towards machine learning based detection and prediction of breakdowns. The authors separately applied deep learning on two available data types (event and trend data) to predict breakdowns. In so doing, they were able to predict breakdowns 20ms in advance with good accuracy. In addition, they utilised explainable AI on these models to elucidate the physics of a breakdown. This pointed them towards an increased pressure in the vacuum system before a breakdown, which they indicated as an option for an improved interlocking system. Their analysis of event data alone also reveals the possibility to predict breakdowns with good accuracy, if there has already been a breakdown in the previous minute, i.e. prediction of follow-up breakdowns.

Previous work within our organisation also informed the present studies. Another dataset from XBOX cavity testing was analysed for missed breakdowns using principal component analysis and neural networks [8]. Very high classification accuracy was reported, but there was suspected duplication of traces in the dataset.

METHODOLOGY

CLARA

Here we use data gathered during the RF conditioning of CLARA's 10 Hz photoinjector (Gun-10), which includes both the RF pulse traces themselves and other non-RF, such as the temperature and pressure inside Gun-10. The RF trace data was gathered before ML was taken into consideration and was therefore not ideal for our purposes, but it was deemed to be sufficient for progress to be made. The trace data was only recorded when the RF breakdown detector was activated and a BD identified, then the conditioning script

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-WEPV021

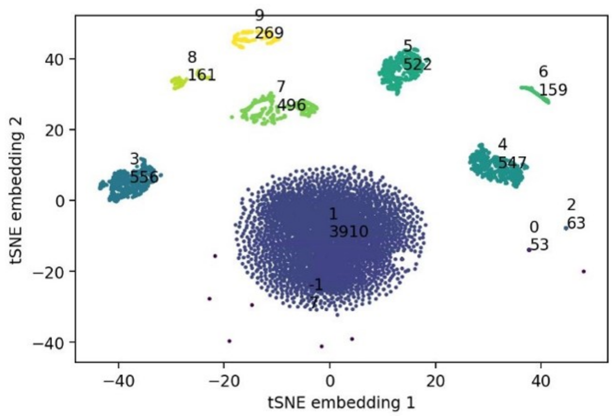

Figure 1: A plot of the results of applying the t-SNE/DB-SCAN method to label the CLARA Gun-10 data set. Each cluster is given an index with its population displayed below, i.e. the cluster in the top right corner of the plot has an index of 6 and a population of 159. As examples of the principal trace types, the large central cluster indexed as 1 represents noise traces, cluster 4 contains only healthy traces, 8 breakdown traces, and 7 is composite (healthy/BD).

would record the pulse associated with the BD, as well as the two previous and two subsequent pulses. Altogether, there were 40 traces recorded per breakdown event (8 traces for each of the 5 pulses). Specifically the traces were: klystron (forward, reverse power, and phase), and cavity,(forward, reverse power, and phase).

In order to provide the ground truth and label each recorded trace as either, noise, healthy, BD or anomaly, the traces were first grouped together by using sklearn's t-SNE[9] (t-distributed Stochastic Neighbour Embedding) and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) functions. The traces from each delivered cluster were then over-plotted and inspected by eye with any pure groups receiving the appropriate label and composite clusters undergoing further t-SNE/DBSCAN analysis until only pure groups remained. Figure 1 shows an example of the clustering that was returned by the t-SNE/DBSCAN method for this data set.

The next step was to construct the βCVAE at the core of the BD detector and at the beginning of the planned BD predictor. After much experimentation with spectrograms, phase traces, and temperature and pressure data, it was found that the most effective input for the βCVAE was a 2D array comprised of the four normalised power traces, with dimensions of 4×1017. The most optimal structure of the found for βCVAE is displayed and described in Fig. 2.

The βCVAE was trained on 4659 healthy traces in order to create the anomaly detector, which was then validated using another 1164 healthy traces. For both training and validation the Adam optimiser and categorical cross entropy loss function were used. Testing the βCVAE involved exposing the algorithm to 706 healthy and 706 BD traces (1412 traces in total) and subtracting the reconstructed traces from the original traces to produce 1D reconstruction error traces. These were then used as an input for a simple dense neural network classifier with one ReLU activated hidden layer with the same dimensions as the input layer and a binary (healthy or BD) softmax activated output layer. Again the Adam optimiser and categorical cross entropy loss function were used.

A confusion matrix was then constructed by comparing the class assigned by the model to the ground truth in order to check for the accuracy and recall of the BD row for the system, as in Table 2. For this dataset, the key statistic was the recall of the BD row, since the accelerator not reacting to a false negative could be damaging to the accelerating structure, whereas reacting to a false positive merely results in a slight reduction in the time efficiency of the accelerator. An accuracy of 96.9% and a BD row recall of 98.0% was achieved using the methods outlined above. We also noted that approximately half of the false negative traces were in fact healthy or anomalous after manual inspection. This is not surprising since the labelling process relied on unsupervised ML processes and, had time allowed, all traces would have been labelled individually by an RF expert.

Future work will include the integration of the ML BD detector into the next version upgrade of the RF conditioning code and the construction of the LSTM RNN for the predictive system. However, before this can be effective, more appropriate data may need to be gathered from CLARA, particularly more traces before a BD event and at a higher RF repetition rate than 10 Hz. Since CLARA RF power is pulsed if we were to follow the methodology set out in Kates-Harbeck et al.[2] we can think of the noise between pulses as missing data . In order to quantify the proportion of data that is effectively missing, or pseudo-missing, we can use the following relation,

$$R_{\mathrm{DATA}} = 1 - D, \tag{1}$$

where R_{DATA} is the proportion of the data that is pseudo-missing and D, the compliment of R_{DATA}, is the dimensionless duty cycle of the RF system, defined as,

$$D = \mathrm{PRR} \times \tau_{\mathrm{pulse}}, \tag{2}$$

where PRR is the pulse repetition rate in Hz and τ_{pulse} is the RF pulse length in seconds. For CLARA's Gun-10 we have a PRR of 10 Hz and an operational pulse length of 2.5 µs, which gives a duty cycle of 2.5×10^{-5}, or 0.0025% and consequently a proportion of pseudo-missing data of 0.999975, or 99.997%. It seems reasonable to assume that for any given system there may exist a "duty cycle threshold" below which BD prediction using ML with RF trace data is not practically possible and which may only be determined experimentally.

CERN

The CERN XBOX data consisted of two primary data types: event and trend. The trend data contained environmental data concerning the test cavity (i.e. temperature,

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-WEPV021

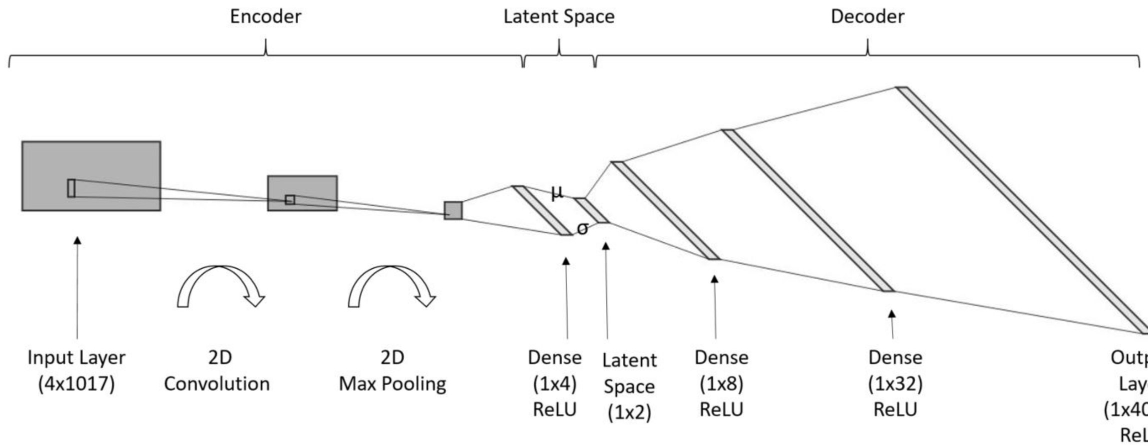

Figure 2: The final structure of the βCVAE for the CLARA dataset study. The encoding half of the βCVAE was constructed by using 2D arrays consisting of the four power traces (4×1017) as inputs to the first 2D ReLU activated convolutional layer with subsequent 2D max pooling layer, the arrays were then flattened and inputted into a 1×4 ReLU activated dense layer before being passed into the 1×2 latent space. The decoder consisted of two dense ReLU activated layers of dimensions 1×8 and 1×32, before a sigmoid activated output layer of 1×4068 ($= 4 \times 1017$). The outputted 1D array was then reshaped into a 4×1017 2D array in order to produce the reconstructed traces with the original input layer dimensions.

vacuum pressure) while the event data contained the signal traces from a number of components of the RF system.

To keep consistency with the CLARA breakdown detector, the trend data was discarded and only the RF traces were used. These traces consisted of 16 channels, of which we excluded 7, as follows. Two of the channels ('DC UP' and 'DC DOWN') corresponded to the Faraday cups upstream and downstream of the RF cavity, and these were used for automated labelling of the samples. A reading of less than -0.05 from either cup would mark that trace as a breakdown. The remaining 5 excluded channels were removed as they were either repeated signals (i.e. the 'PSR log' channel repeated 'PSR Amplitude' channel, but with log scaling) or essentially noise (such as the Beam Loss Monitor signal, which was indistinguishable from noise).

Further filtering was applied to ensure the quality of the data, and traces were removed wherein the mean and variance of the amplitude traces indicated that the RF cavity was not active ($\bar{x} \times \sigma^2(x) < 1e{-}4$), and the signals were therefore considered noise. This filtering and classification resulted in 254,656 samples, of which 5,930 samples contained a breakdown.

We then developed a beta variational autoencoder model, as shown in Fig. 3, to reconstruct only healthy signals. The network was trained using 90% of the healthy signals. The Adam optimiser was used with a learning rate of $1e-3$, with the error function defined in equation (3). The network was trained to convergence, which took 27 epochs. For β a value of 5 was chosen by grid search.

$$E(x) = \beta D_{kl}(N(0,1)||P(x|\mu,\sigma)) + abs(x - \bar{x}) \quad (3)$$

Note that D_{kl} represents the Kullback-Liebler divergence[10].

Once sufficiently trained to reconstruct healthy signals, we utilised this *overfitting* to detect breakdown events as anomalies. That is to say, when the network fails to reconstruct the signal well, we can be reasonably assured that this represents a deviation from normal operation and thus a breakdown event.

In order to classify the breakdown events, we began by passing all breakdown events and an equal number of non-breakdown events through the autoencoder, taking the reconstruction error for each channel and compiling those values into a vector. Applying a K-nearest neighbour algorithm to the per-channel reconstruction error vector resulted in mediocre performance, as shown in Table 1. As such, we elected to implement a simple multi-layer neural network to perform the classification, while also concatenating the latent space representation of the example to the per-channel reconstruction errors to form the input vector. This new neural network was then trained to convergence in 20 epochs using the Adam optimiser, with a learning rate of $1e-2$ and a binary cross-entropy loss function. Results from this network are shown in Table 3.

Table 1: kNN Results on CERN XBOX Data

	Positive	Negative
True	89.6%	98.8%
False	10.4%	1.2%

Early work into prediction of breakdowns was undertaken by replacing the classification network with an LSTM which was trained on a label of time-until-breakdown. Unfortunately, this achieved little success with our most effective attempt achieving 73% accuracy at predicting breakdowns within 30 second windows. We hypothesise this is due to the sparse nature of the sampling of shots, and future work

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-WEPV021

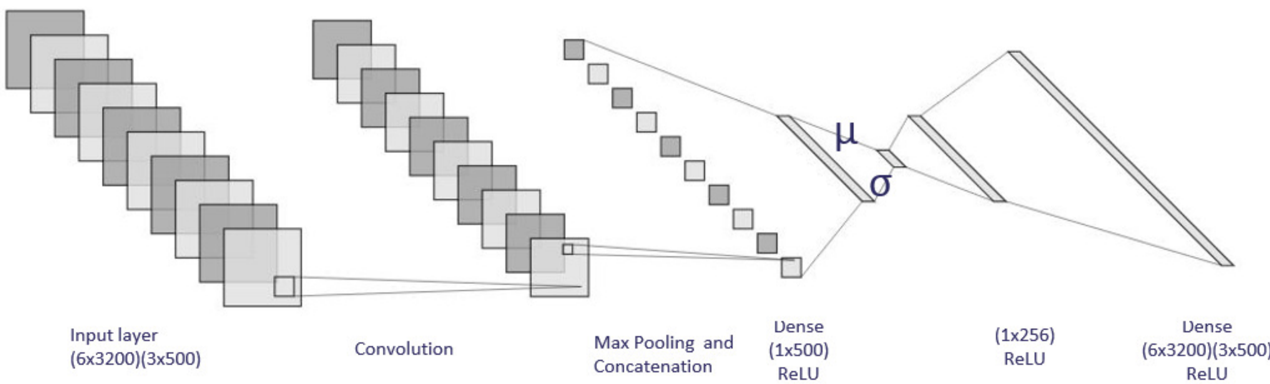

Figure 3: For the CERN XBOX data, three of the input channels were of a lower dimension, and each channel was processed by a CNN with filters of size 1×5, with the resulting feature vectors flattened and concatenated before being passed to a dense ReLu layer and encoded into the latent space. The decoder is constructed of a single dense ReLu layer with each channel consisting of a linear dense layer.

will involve collecting a higher frequency sampling of shots such that this prediction might be enabled.

RESULTS

As can clearly be seen, both networks produced strong results very high recall and accuracy. In particular, the high recall value is significant for such an unbalanced dataset. If of a system of this type were to be deployed to an edge ML system on an accelerator, a low false positive rate would be extremely important for operator trust of the system.

Table 2: Results on CLARA Data

	Positive	Negative
True	95.8%	98.0%
False	4.2%	2.0%

Table 3: Results on CERN XBOX Data

	Positive	Negative
True	97.9%	99.6%
False	2.1%	0.4%

It is of note that approximately half of the false positives in the CERN XBOX data are in fact true positives that were mislabelled by the automated labelling, as verified by manual inspection.

CONCLUSION

We find that the application of a variational autoencoder as an anomaly detector is extremely effective as a breakdown detector for RF cavities. A significant benefit of this approach is that it requires only healthy signals to train a strong detector, in contrast to supervised approaches which require careful balancing of positive and negative signals. Obermair et al.[7] showed that environmental trend data was sufficient to predict breakdowns with good accuracy up to 20ms in advance. They also found that shot traces are sufficient to predict breakdowns with good accuracy, if and only if there has already been a breakdown in the last minute. In agreement with their results, we find that the shot traces from the cavity alone are not sufficient at this sampling frequency to predict breakdowns ex nihilo. We plan to collect additional data from CLARA with complete shot capture. With this data, we hope to demonstrate breakdown prediction using only per-shot phase and amplitude traces and the anomaly detection method presented herein.

ACKNOWLEDGEMENTS

The authors would like to thank the CLIC team and the Paul Scherrer Institute for the provision of their high-gradient structure test data, in addition to Christoph Obermair and Lee Millar from CERN for their advice on interpreting it. We would also like to thank Hannah Kockelbergh (The University of Liverpool) for the studies that led on to this work, as well as STFC's Scientific Machine Learning (SciML) Group, particularly Keith Butler, for supervising A. J. Gilfellon's work on this project.

REFERENCES

[1] I. Higgins *et al.*, "Beta-vae: Learning basic visual concepts with a constrained variational framework," 2016.

[2] J. Kates-Harbeck, A. Svyatkovskiy, and W. Tang, "Predicting disruptive instabilities in controlled fusion plasmas through deep learning," *Nature*, vol. 568, no. 7753, pp. 526–531, Apr. 2019. DOI: 10.1038/s41586-019-1116-4. https://doi.org/10.1038/s41586-019-1116-4

[3] P. Craievich *et al.*, "Consolidation and extension of the high-gradient linac RF technology at PSI," *Proc. LINAC'18*, pp. 937–940, 2019.

[4] R. Zennaro *et al.*, "High power tests of a prototype x-band accelerating structure for CLIC," *Proc. IPAC'17*, pp. 4318–4320, 2017.

[5] A. Solopova *et al.*, "SRF cavity fault classification using machine learning at CEBAF," in *10th Int. Particle Accelerator Conf.(IPAC'19), Melbourne, Australia, 19-24 May 2019*,

ISBN: 978-3-95450-221-9 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2021-WEPV021

JACOW Publishing, Geneva, Switzerland, 2019, pp. 1167–1170.

[6] C. Tennant, A. Carpenter, T. Powers, A. Shabalina Solopova, L. Vidyaratne, and K. Iftekharuddin, "Superconducting radio-frequency cavity fault classification using machine learning at Jefferson Laboratory," *Physical Review Accelerators and Beams*, vol. 23, no. 11, Nov. 2020, ISSN: 2469-9888. DOI: 10.1103/physrevaccelbeams.23.114601. http://dx.doi.org/10.1103/PhysRevAccelBeams.23.114601

[7] C. Obermair *et al.*, "Machine learning models for breakdown prediction in RF cavities for accelerators," *Proc. IPAC'21*, 2021.

[8] H. Kockelbergh and D. Dunning, "Summary of RF conditioning data analysis for cavity T24N5_L1 on Xbox 3," *STFC technical note*, 2020.

[9] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE.," *Journal of machine learning research*, vol. 9, no. 11, 2008.

[10] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.