ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOMS054

DATA AUGMENTATION FOR BREAKDOWN PREDICTION IN CLIC RF CAVITIES

H. S. Bovbjerg *[1,2], C. Obermair[1,3], A. Apollonio[1], T. Cartier-Michaud[1], W. Millar[1], Z.H. Tan[2], M. Shen[2], D. Wollmann[1]
[1]CERN, Geneva, Switzerland [2]Aalborg University, Aalborg, Denmark [3]TU Graz, Graz, Austria

Abstract

One of the primary limitations on the achievable accelerating gradient in normal-conducting accelerator cavities is the occurrence of vacuum arcs, also known as RF breakdowns. A recent study on experimental data from the CLIC XBOX2 test stand at CERN proposes the use of supervised machine learning methods for predicting RF breakdowns. As RF breakdowns occur relatively infrequently during operation, the majority of the data was instead comprised of non-breakdown pulses. This phenomenon is known in the field of machine learning as class imbalance and is problematic for the training of the models. This paper proposes the use of data augmentation methods to generate synthetic data to counteract this problem. Different data augmentation methods like random transformations and pattern mixing are applied to the experimental data from the XBOX2 test stand, and their efficiency is compared.

INTRODUCTION

The RF cavities of the Compact LInear Collider (CLIC) are designed to operate at a gradient of ~100 MV/m [1]. One of the primary limitations on the achievable gradient in normal conducting RF cavities is the occurrence of RF breakdowns, which can degrade a passing beam and potentially result in damage to the cavity surface [2–4]. In order to minimize the impact of breakdowns during the cavity commissioning and operation, CERN's CLIC test stands [5] employ an automatic conditioning algorithm [6, 7]. The algorithm monitors how frequently breakdowns occur during operation and dynamically adjusts the gradient based on a preset breakdown-rate threshold [8]. In this approach, the handling of breakdowns is therefore purely reactive, thus breakdowns cannot be prevented beforehand.

In a recent study, a deep learning approach was proposed with the goal of (1) performing data-driven breakdown investigation and (2) studying the possibility of adopting a predictive conditioning algorithm. The study was based on historical data of the CERN XBOX2 test stand, consisting of 124 505 healthy RF pulses and 479 breakdown events [9].

Previously, it has been noted that breakdowns occur predominantly in groups as opposed to isolated, single events. This observation has led to the classification of breakdown events as either *primary breakdowns*, which are purely stochastic, and *followup breakdowns*, which are thought to be a consequence of the previous breakdown [10]. Using the XBOX2 data, neural networks were able to predict the occurrence of followup breakdowns. However, the prediction accuracy varied depending on different data used for the prediction, e.g. for different adopted parameters for cavity powering. This variation indicates that the models were not able to generalize well to unseen data [11]. Specifically, the bad generalization is due to the low number of breakdown events compared to the number of healthy events, i.e. the so-called high *class imbalance*. We therefore investigated the use of time series data augmentation methods for improving the generalization capabilities of CLIC breakdown prediction. The basic principle of these methods involves generating synthetic patterns that resemble real data to better represent the underlying distribution of the underrepresented class in the data set. This is an established practice for image recognition tasks [11–13] and is also used for speech and audio [14, 15].

The paper is structured as follows: first, a summary of the prior work is given, including a description of the data and model used in our study. Next, an overview of the augmentation methods used in this paper is presented. Finally, the conducted experiments are described, and their results are discussed.

PRIOR WORK

This section summarizes the prior work which this work builds upon, including a description of the data set used in the study, and a description of the RF breakdown prediction models used.

XBOX2 Data Set

The XBOX2 test stand is one of three experiments used to test the prototype 12GHz RF components for the CLIC project at CERN. Fundamentally, the test stand is composed of a 50 MW klystron, pulse compressor, and high-power RF load. A more detailed description of the setup is available elsewhere [7, 9].

In 2018 this test stand produced 90 GB of data during an operational period of six months, consisting of so-called trend and event data [9]. The trend data contains 30 different scalar values such as temperatures and pressures measured at different locations in the test-stand. The event data consists of time-series measurements of the RF signals at different locations in the waveguide network and the current detected by two Faraday cups. A summary of the data is given in Fig. 1. Here, two features of the forward travelling wave signal F2 (see fig. 2), namely the maximum (blue) and the pulse width (green), are shown with respect to the RF cavity pulses. Additionally, the cumulative breakdown count (red)

* holger.severin.bovbjerg@cern.ch

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOMS054

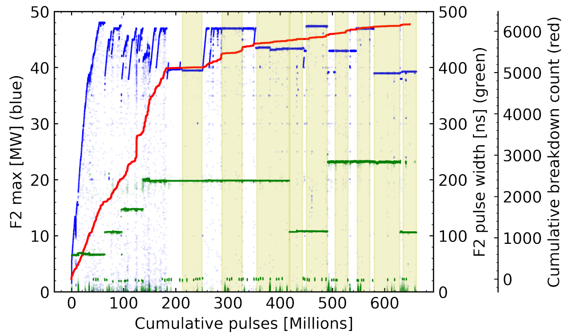

Figure 1: Overview of the conditioning period, of all data analyzed [9]. It shows the maximum of the power amplitude of the forward travelling wave signal F2 (blue), its pulse width (green), and the cumulative breakdown count (red).

is plotted. The yellow area represents the periods with constant operational settings used for further analysis, leading to a total of 479 breakdowns and 124 505 healthy RF pulses, as not every pulse is stored. Given the previously observed probabilistic behavior of breakdowns, the data is further divided into 229 primary and 250 followup breakdowns. Primary breakdowns were defined as not having occurred within 3000 pulses of the previous breakdown, corresponding to one minute of operation in the test stand, which has a pulse repetition rate of 50 Hz.

Modelling of RF Breakdowns

In [9], a number of neural network architectures were investigated to predict breakdowns using trend data and event data. These two experiments were further split into the prediction of primary breakdowns and followup breakdowns. Formally, the prediction of breakdowns is defined as finding a model $f(\cdot)$ that uses the observed data $\mathbf{x}_i$ to predict the label (healthy or breakdown) of the next time stamp y_{i+1}, where i is the current time step.

The model performance is measured using the Area under the Receiver operating characteristics curve (AR) [16]. This score is defined as the probability that a model will classify a randomly selected breakdown event as more likely to be a breakdown than a randomly selected healthy event. An AR score of 100% means that the model is able to perfectly predict the class labels, and a score of 0% corresponds to a classifier which predicts all labels wrong.

Primary breakdowns proved to be difficult to predict with available event data, whereas it was possible to predict followup breakdowns with an AR score of up to 89.7% ± 8.1%. We aim at further improving the Fully Convolutional Network (FCN), achieving this result, with data augmentation.

DATA AUGMENTATION

The XBOX2 data consist of a number of time series, therefore we focus on time series augmentation methods, which can generally be divided into four categories: random transformations, pattern mixing, generative models and decomposition models [11]. In this study, we only consider random transformation methods and pattern mixing. We do not consider generative models due to the computational cost and their high number of parameters. Furthermore, due to the non-periodic nature of the XBOX2 data, decomposition models are deemed inapplicable. Illustrations of all applied methods are seen in Figure 2.

Random Transformation

Random transformation methods apply different types of transformations to the data, in order to generate new synthetic samples. Random transformation methods assume that the transformations are representative of the data characteristics [11], i.e. they can be introduced without changing the fundamental nature of the signals. Typically, augmentation methods alter the values, the time steps or the frequencies in a signal, i.e. transformations take place in the magnitude, time, or frequency domain. In the case of the XBOX2 data, frequency transformations are not applicable, as the data is not periodic.

A simple random transformation method in the magnitude domain is *noise addition*, also known as *jittering*. Here, a noise vector $\boldsymbol{\alpha}$ is sampled from a zero mean Gaussian $\sim \mathcal{N}(0, \sigma^2)$, which is then added to a data sample $\mathbf{x}$ to generate a synthetic sample $\mathbf{x}'$ such that $\mathbf{x}' = \mathbf{x} + \boldsymbol{\alpha}$. Adding noise has been shown to improve generalization of neural networks [17].

Another similar strategy, known as *magnitude scaling* [18], scales the data sample by a Gaussian scaling vector $\boldsymbol{\beta} \sim \mathcal{N}(1, \sigma^2)$, such that $\mathbf{x}' = \mathbf{x} \cdot \boldsymbol{\beta}$. A more advanced version of this approach is known as *magnitude warping* [18]. Here the scaling vector is based on interpolation from a cubic spline S with k knots, with the knots being drawn from a Gaussian $\sim \mathcal{N}(1, \sigma^2)$.

Random transformation methods that act in the time domain include warping and slicing methods. *Window slicing* generates new samples by only selecting a certain percentage W of the available samples, and interpolating back to the original number of samples. *Warping* in time involves perturbing the individual data point of a sample in time. Given a warping function τ, defined by a cubic spline S with k knots drawn from a Gaussian distribution $\sim \mathcal{N}(1, \sigma^2)$, a new sample is found as $\mathbf{x}' = x_{\tau(1)}, \ldots, x_{\tau(t)}, \ldots, x_{\tau(T)}$, with T being the sample length.

Pattern Mixing

Pattern mixing techniques seek to generate synthetic samples by mixing features of multiple data samples. In its simplest form, pattern mixing takes the mean between two or more signals of the same class. However, this method might remove distinguishing features from the signal, due to smoothening from the mean operator.

A popular method for pattern mixing is known as Synthetic Minority Oversampling Technique (*SMOTE*) [19]. The SMOTE method takes a sample of the minority class $\mathbf{x}$ and randomly selects a k-nearest neighbor $\mathbf{x}_{\mathbf{NN}}$. The absolute difference between them is then found and scaled by a random scalar $\lambda \sim \mathcal{U}(0, 1)$, and the new sample is found as $\mathbf{x}' = \mathbf{x} + \lambda|\mathbf{x} - \mathbf{x}_{\mathbf{NN}}|$.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOMS054

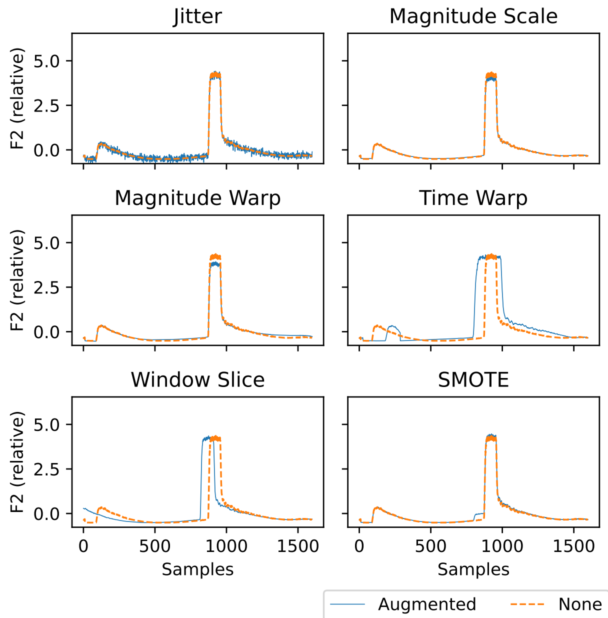

Figure 2: Illustration of augmentation techniques applied to a forward travelling wave signal F2 from the XBOX2 data set.

EXPERIMENTS

To test whether data augmentation is beneficial for the prediction of RF breakdowns, a series of simulation experiments have been carried out. For each of the selected data augmentation methods, we train the FCN model following the approach of [9]. Each augmentation method includes a number of hyperparameters, which we choose based on recommended values from literature [11].

In all our data augmentation methods, we oversample the minority class and take only 2.5% of healthy events, i.e. 3113 events, similarly to prior work [9]. Considering the whole data set, we augment 3113 healthy and all 250 followup breakdowns, to acquire 3113 healthy and 3113 followup breakdowns. Data augmentation aims to remove the class imbalance, making class weighting used in previous work [9], not always necessary. The best results of each method are summarized in Table 1. Methods with class weighting are marked with (*).

To fairly assess the model performance with data augmentation, a *train on synthetic test on real* paradigm is used. This means that the models are trained on a training set containing synthetic data, however, the validation and test set is kept untouched. In this manner, the performance using data augmentation can directly be compared to the baseline model trained without data augmentation.

The periods of stable operation are used for k-fold cross-validation. This means that one group is set aside as a validation set, using the rest for training. Each stable operation is used as a validation set once. The mean AR score is then reported as AR_μ with standard deviation AR_σ. After fine-tuning manual model parameters, the model is finally trained on both training and validation set, and tested on an unseen stable operation period with a performance AR_t.

Table 1: Best AR Scores for Various Implemented Augmentation Methods

Augmentation	AR_μ [%]	AR_σ [%]	AR_t [%]
None*	89.7	8.1	91.1
Jitter	90.0	2.9	84.2
Magnitude Scaling	89.4	5.1	87.6
Magnitude Warp	90.4	4.9	86.0
Time Warp	90.8	4.6	84.6
Window Slicing*	89.4	5.0	90.4
SMOTE*	91.0	5.2	89.8

RESULTS & DISCUSSION

In Table 1, we present the results obtained from applying data augmentation methods to the XBOX2 data when predicting followup breakdowns. When comparing the results using no data augmentation to the results with data augmentation, a slight increase in the mean AR score on the validation set is seen for jittering, magnitude warping, time warping and SMOTE. Magnitude scaling and window slicing instead show a slight decrease. The SMOTE method achieves the largest improvement over no data augmentation and yields an improvement of 1.3%, when keeping the class weighting from the previous study. The best result achieved with no class weighting was for time warping, with an improvement of 1.1%.

Looking at the standard deviation, all augmentation methods yield a significant decrease. This means that the performance of the trained model varies less on different validation sets when using data augmentation, and that the models are able to generalizes better, independently of the stable operation period. The best performance, with respect to AR_σ, was achieved by jittering which decreased the standard deviation by 5.2% compared to no augmentation, yielding a standard deviation of 2.9%. Note, that AR_t is only used to validate the model's generalization capabilities by testing whether AR_t is within $AR_\mu \pm 2AR_\sigma$.

CONCLUSION

In this paper, we investigated different techniques to improve existing RF breakdown prediction models through the use of data augmentation methods applied to time series data from CERN's XBOX2 test stand. We conclude that data augmentation improves the standard deviation of our model independent of the technique, making the used model more robust and generic. The performance of the model, however, only improve slightly dependent on the technique. The best performance was achieved using the SMOTE method, keeping the class weighting from the original study. SMOTE improved the average model performance by 1.3% and decreased the standard deviation by 2.9%. The achieved results provide new insights for the development of a proactive and dynamic conditioning algorithm for CLIC RF cavities.

ISBN: 978-3-95450-227-1 ISSN: 2673-5490 doi:10.18429/JACoW-IPAC2022-TUPOMS054

REFERENCES

[1] CLICdp Collaboration, *The Compact Linear Collider (CLIC) - 2018 Summary Report*, P. Burrows, Ed. 2018, vol. 2, doi: 10.23731/CYRM-2018-002

[2] H. Wiedemann, *Particle Acelerator Physics*, 4th ed. Springer, 2015.

[3] B. Woolley *et al.*, "High-gradient behavior of a dipole-mode rf structure," *Physical Review Accelerators and Beams*, vol. 23, 2020, doi:10.1103/PhysRevAccelBeams.23.122002

[4] A. Hassanein *et al.*, "Effects of surface damage on rf cavity operation," *Phys. Rev. ST Accel. Beams*, vol. 9, p. 062 001, 6 2006, doi:10.1103/PhysRevSTAB.9.062001

[5] T. Lucas *et al.*, *High power testing of a prototype clic structure: Td26cc r05 n3*, 2018.

[6] W. Wuensch *et al.*, "Experience Operating an X-band High-Power Test Stand at CERN," in *Proceedings of the 5th Int. Particle Accelerator Conf.*, 2014, doi:10.18429/JACOW-IPAC2014-WEPME016

[7] L. Millar, "Operation of Multiple Accelerating Structures in an X-Band High-Gradient Test Stand," Presented 22 Jul 2021, 2021, https://cds.cern.ch/record/2798232

[8] N. Catalan-Lasheras *et al.*, "Commissioning of XBox-3: A Very High Capacity X-band Test Stand," in *Proc. LINAC'16*, East Lansing, MI, USA, Sep. 2016, pp. 568–571, doi:10.18429/JACoW-LINAC2016-TUPLR047

[9] C. Obermair *et al.*, "Explainable machine learning for breakdown prediction in high gradient rf cavities," 2022, doi: 10.48550/ARXIV.2202.05610

[10] W. Wuensch *et al.*, "Statistics of vacuum breakdown in the high-gradient and low-rate regime," *Phys. Rev. Accel. Beams*, vol. 20, p. 011 007, 1 2017, doi:10.1103/PhysRevAccelBeams.20.011007

[11] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *PLOS ONE*, vol. 16, no. 7, e0254841, 2021, doi: 10.1371/journal.pone.0254841

[12] Q. Wen *et al.*, "Time series data augmentation for deep learning: A survey," 2021, doi:10.24963/ijcai.2021/631

[13] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, 2014, doi:10.48550/ARXIV.1409.1556

[14] T. Ko, V. Peddinti, D. Povey, and S. Khudanpur, "Audio augmentation for speech recognition," in *INTERSPEECH*, 2015.

[15] W. Hartmann, T. Ng, R. Hsiao, S. Tsakalidis, and R. M. Schwartz, "Two-stage data augmentation for low-resourced speech recognition," in *INTERSPEECH*, 2016.

[16] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010

[17] G. An, "The Effects of Adding Noise During Backpropagation Training on a Generalization Performance," *Neural Computation*, vol. 8, no. 3, pp. 643–674, 1996, doi:10.1162/neco.1996.8.3.643

[18] T. T. Um *et al.*, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," 2017, doi:10.1145/3136755.3136817

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953