frontiers | Frontiers in Physics

TYPE Brief Research Report
PUBLISHED 30 May 2025
DOI 10.3389/fphy.2025.1553993

OPEN ACCESS

EDITED BY
Didier Mazon,
CEA Cadarache, France

REVIEWED BY
Simone Spampinati,
National Laboratory of Frascati (INFN), Italy
Á. Sánchez-Villar,
Princeton Plasma Physics Laboratory (DOE), United States

*CORRESPONDENCE
Lynda Boukela,
lynda.boukela@desy.de
Annika Eichler,
annika.eichler@desy.de

RECEIVED 31 December 2024
ACCEPTED 05 May 2025
PUBLISHED 30 May 2025

CITATION
Boukela L, Branlard J and Eichler A (2025) Exploring NAS for anomaly detection in superconducting cavities of particle accelerators.
Front. Phys. 13:1553993.
doi: 10.3389/fphy.2025.1553993

Exploring NAS for anomaly detection in superconducting cavities of particle accelerators

Lynda Boukela[1]*, Julien Branlard[1] and Annika Eichler[1,2]*

[1]Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany, [2]Hamburg University of Technology, Institute of Control Systems, Hamburg, Germany

The European X-Ray Free Electron Laser is the largest particle accelerator for X-ray laser generation worldwide. To ensure a safe and efficient operation, the plant uses various monitoring systems, especially in the linear accelerator. The low-level radio frequency system has shown reliability in diagnostics, particularly in quench detection. A quench refers to a superconducting radio frequency cavity losing its superconductivity and possibly causing a downtime. The diagnostics solution, however, can be enhanced in terms of robustness and functionality. Currently, the focus is on integrating artificial intelligence to improve quench identification. Thus, a *lightweight* machine learning-assisted approach targeting FPGA deployment is developed. It relies on the augmentation of a physical model-based anomaly detection approach with neural network models to distinguish the quenches from the other anomalies. This paper presents the solution in which neural architecture search is applied, and elaborates on how visualizing and analyzing the anomaly detection results can provide critical insights for both short-term diagnostics and long-term pattern identification.

KEYWORDS

anomaly detection, particle accelerators, neural architecture search, data visualization, superconductivity

1 Introduction

The European X-Ray Free Electron Laser (EuXFEL) is the largest particle accelerator for X-ray laser generation worldwide. It spans over 3.4 km in Hamburg, Germany, and serves several hundred users each year. The interdisciplinary researchers benefit from an extremely intense laser light generated at a rate of 27,000 flashes per second, with an electron acceleration reaching high energies of up to 17 GeV. The linear accelerator (linac) achieves acceleration with almost 800 superconducting radio frequency cavities (SRFCs), organized into 25 stations. The SRFCs act as resonators that propel charged particles when operated at their resonance frequency of 1.3 GHz. They are currently operated in a pulsed mode with a pulse repetition rate of 10 Hz. An overview of the accelerator is shown in Figure 1.

The safe and efficient operation of large and complex facilities, such as the EuXFEL, is crucial for successful experiments by the users. Anomalous behaviors are encountered daily in different areas of the plant. Thus, various monitoring systems were deployed during commissioning. At the linac, the low-level radio frequency (LLRF) system controls and monitors the SRFCs. It ensures a stable accelerating field, averaging 23.6 MV/m, additionally, it diagnoses the radio frequency (RF) signals and other measurements to

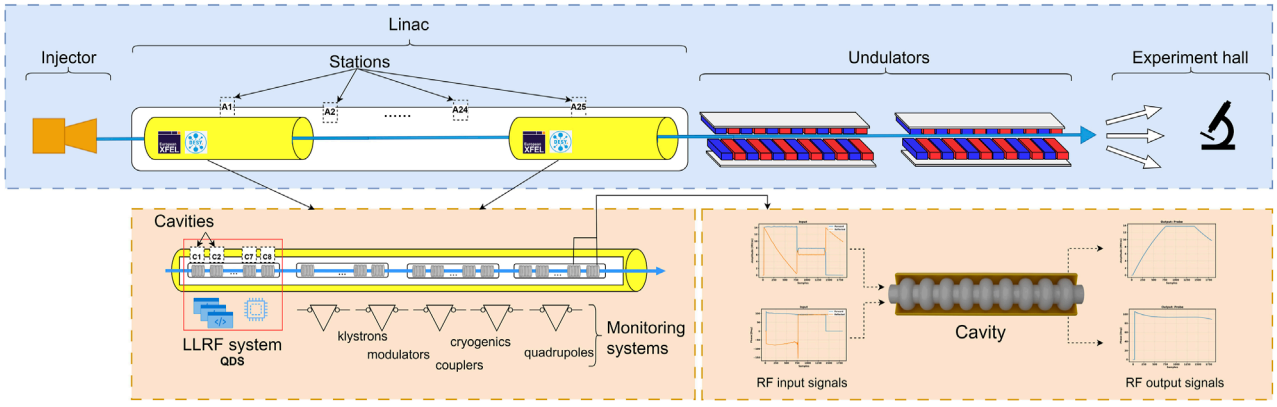

FIGURE 1
Overview of the EuXFEL with a focus on the SRFCs location within the stations and the linac.

report any anomaly and to trigger countermeasures accordingly. However, under certain conditions, these systems can be inflexible and suffer from robustness issues. Therefore, upgrades and improvements have been continuously conducted, with a recent growing focus on artificial intelligence (AI) techniques. These techniques have also been explored in other facilities for anomaly classification and prediction. For example, quenches are classified through decision trees and neural networks at the relativistic heavy ion collider (RHIC) [1]. At the continuous electron beam accelerator facility (CEBAF), a long short-term memory-convolutional neural network is developed to predict faults in the accelerating cavities [2].

Downtime can occur when an SRFC loses its superconductivity due to a quench. When an SRFC exceeds its maximum sustainable gradient, it quenches and transitions to a normal conducting state. This translates to a loss of the gradient and a drop in the quality factor (Q_L), where Q_L is an indicator of the field coupling and power dissipation in the cavities. The currently deployed quench detection system (QDS) relies on a statistical analysis of the Q_L [3]. While the QDS is effective in detecting quenches, it also generates a considerable number of false alarms, triggering with faults different from quenches [4]. A machine learning (ML)-enhanced solution has therefore been explored, to improve anomaly detection and categorization. For general anomaly detection, the method relies on the SRFC model that couples the electromagnetic and mechanical dynamics. It employs the non-linear parity space-based method and the generalized likelihood ratio (GLR) [4, 5]. To identify the anomalies, the system can be augmented with ML techniques. The k-medoids clustering algorithm, using the Euclidean (EUC) and the Dynamic Time Warping (DTW) similarity measures, has been explored to identify quenches [6]. This paper continues the previous work by presenting a neural network-based approach. Various multilayer perceptrons (MLPs) [7] are used to learn enhanced decision boundaries in terms of separation of the quenches from the other anomalies in the distance space of the clustering medoids (actual-data centroids). The MLP architectures are learned through two approaches, a handcrafted method with varying neuron counts per layer and an optimization-based approach using neural architecture search (NAS). The latter uses the evolutionary optimization to find a *lightweight* model that maximizes the detection performance while minimizing the inference latency, given the goal of firmware implementation. The offline implementation analyzes RF pulse data collected daily, providing both short-term monitoring capabilities and long-term trend analysis. This helps LLRF experts diagnose issues promptly and understand broader patterns. Both the evaluation and exploitation of the solution are detailed and discussed.

2 Preliminaries

The electromagnetic and mechanical dynamics of the SRFCs are modeled with the input forward signal, which corresponds to the RF signal driving the cavity and the output probe signal, which corresponds to the RF signal coupled out from the field inside of the cavity. The influencing parameters are the detuning, which is the delta between the driving and the resonance frequencies, and the half bandwidth, which is an indicator of an SRFC sensitivity towards the detuning [8]. This model has been used to develop a parity space-based method for anomaly detection. The method captures inconsistencies relative to the expected outputs in the form of residuals, which are subsequently evaluated with the GLR. When the latter exceeds a predefined threshold, it indicates a fault occurrence with high probability [4],

$$\begin{cases} \mathrm{GLR}(k) = \dfrac{K}{2}\left(\dfrac{1}{K}\displaystyle\sum_{i=k-K+1}^{k} r(i)^{\top}\right)\sigma^{-1}\left(\dfrac{1}{K}\displaystyle\sum_{i=k-K+1}^{k} r(i)\right), \\ \mathrm{Anomaly}_{\mathrm{GLR}} = \begin{cases} 1, & \text{if } \exists\ k \text{ such that } \mathrm{GLR}(k) > T, \\ 0, & \text{otherwise,} \end{cases} \end{cases}$$

where $r(i)$ is the evaluated and discretized residual, K is the size of the moving evaluation window, σ is the variance of the nominal residual, and T is the predefined threshold typically equal to 10.8, which corresponds to a desired false positive rate (FPR) of 0.0003% (assuming that the GLR is following a χ^2 distribution) and can be tuned empirically. Figure 2 illustrates the behavior of the SRFCs, their waveforms and corresponding GLR, under nominal and quenching conditions.

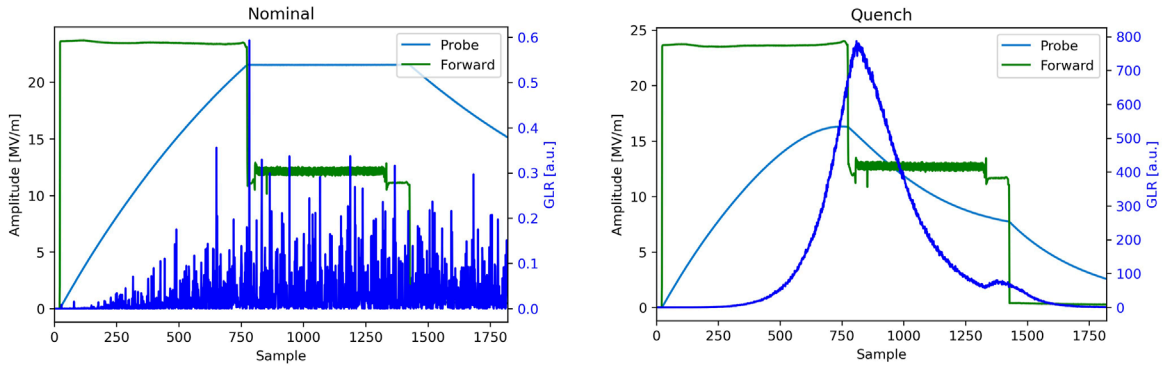

FIGURE 2
Illustrative examples of the SRFC behavior under different conditions. The input forward signal, which corresponds to the RF signal driving the cavity and the output probe signal, corresponding to the RF signal coupled out from the field inside of the cavity. (Left) RF waveforms and GLR obtained under a nominal condition of the SRFC, the GLR is a noisy signal not exceeding the threshold. (Right) RF waveforms and GLR obtained under a quenching condition of an SRFC. This is characterized by a loss in the accelerating gradient (drop of the probe) from approximately 23 MV/m to 17 MV/m. It has been noticed that a bell-shaped GLR corresponds to a quench, with different peak values (approximately equal to 800 in this example), and therefore different center points within the pulse.

The GLR can be augmented with different lightweight ML models in order to distinguish quenches from other anomalies. Preliminary results with a clustering solution based on k-medoids have been obtained [6]. Given $X = \{x_1, x_2, \ldots, x_n\}$, with $x_i \in \mathbb{R}^d$, a dataset consisting of $n = 76$ quench GLR traces of dimension $d = 1819$, the K-medoids [9], which is a clustering algorithm that groups data by choosing actual data representatives, called medoids, as cluster centers, is used to cluster the GLR traces. Two clusters are built for each similarity metric (note that the number of clusters has been noticed through visualization during data preparation), and therefore two medoids, $M_{DTW} = \{m_{DTW1}, m_{DTW2}\}$ and $M_{EUC} = \{m_{EUC1}, m_{EUC2}\}$, have been obtained with DTW and EUC, respectively. The similarity measures are defined as,

$$\mathrm{EUC}(x_1, x_2) = \sqrt{\sum_{i=1}^{d} \left(x_1^i - x_2^i\right)^2},$$

and

$$\mathrm{DTW}(x_1, x_2) = \arg\min_{i,j} \sum \mathrm{dist}\left(x_1^i, x_2^j\right),$$

where $i, j \in \{1, 2, \ldots, d\}$ represent the sample indices of the traces x_1 and x_2. Any distance measure (dist) can be used for DTW, in this case, the Euclidean distance is applied. More details can be found in [6]. With the two similarity measures explored, significant improvements in terms of false alarms have been achieved. However, the decision boundaries used in the distance space could be improved. Therefore, we propose here to use ML to learn a new and efficient separation of the quenches from the rest of anomalies.

3 Methodology

We use an MLP, which is a type of artificial neural network consisting of multiple fully-connected layers of neurons, an input layer, one or more hidden layers, and an output layer. Each neuron in a layer is connected to every neuron in the previous and next layer, making the network fully connected. When provided with a new anomalous GLR trace $g \in \mathbb{R}^d$, in order to make decision whether it is a quench or not, its distance to the quench medoids is computed and fed to the MLP to assess its similarity with the quench clustering model. $\mathrm{MLP}_{\mathrm{EUC}}$ is built to assess the similarity based on EUC and has therefore an input layer with two neurons, the first neuron is fed with $\mathrm{EUC}(g, m_{EUC1})$ (EUC-based similarity between g and the first medoid m_{EUC1}) and the second neuron is fed with $\mathrm{EUC}(g, m_{EUC2})$ (EUC-based similarity between g and the second medoid m_{EUC2}). Similarly, $\mathrm{MLP}_{\mathrm{DTW}}$ takes as inputs $\mathrm{DTW}(g, m_{DTW1})$ and $\mathrm{DTW}(g, m_{DTW2})$ with DTW as similarity measure.

To build the two MLPs, data from 2021 to 2022 have been exploited. A total of 146,811 anomalous GLR traces is used, with a 70% – 30% split between the training and test sets. Each of the two architectures includes up to four hidden layers, this hidden architecture is learned through two approaches, a handcrafted approach and an optimization-based approach [10], with the optimization aiming at learning a lightweight architecture. In the first approach, we explore architectures with different layer structures. Denoting the varied neuron counts per layer by h, we define architectures with three hidden layers where h is doubled or halved in the middle layer. In addition, architectures where h is progressively increased or shrunk by a factor of two have been explored. We have also examined uniform architectures in which the number of neurons h remains consistent across all layers.

The second approach searches for the model hyper-parameters by leveraging NAS with the evolutionary optimization as tuner (see Figure 3). The architecture evaluation is designed as a hardware-agnostic multi-objective optimization. It rewards higher area under the curve of the receiver operating characteristic (AUROC), given the binary classification problem, and it penalizes higher floating point operations per second (FLOPs) in order to enhance the model

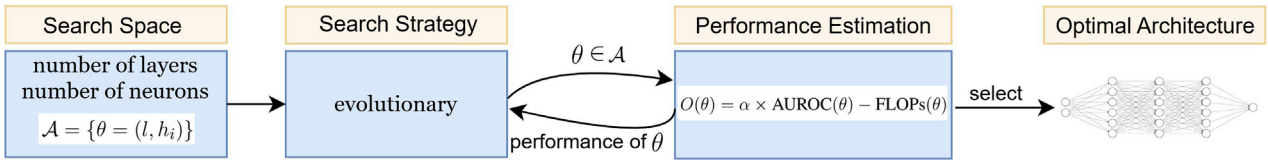

FIGURE 3
Flow of the neural architecture search to learn the MLPs.

efficiency. We therefore define an objective function as a weighted sum of AUROC and normalized FLOPs,

$$\begin{cases} \theta^* = \arg\max_{\theta \in \mathcal{A}} O(\theta), \\ O(\theta) = \alpha \times \text{AUROC}(\theta) - (1-\alpha)\ \text{FLOPs}(\theta), \end{cases}$$

where, $\mathcal{A} = \{\theta = (l, h_i) \mid l, h_i \in \mathbb{N},\ l \in [l_{\min}, l_{\max}],\ h_i \in [h_{\min}, h_{\max}]$ with $i \in \{1, .., l\}\}$ is the search space with the set of all possible architectures constrained by the number of hidden layers (l) and number of neurons (h_i) in layer i. The minimum number of layers was set to $l_{\min} = 1$, and the maximum number of layers was set to $l_{\max} = 3$, as discussions are ongoing with the firmware experts to define what a lightweight model is, we would like to keep the maximum number of hidden layers as small as possible. α is a scaling factor that controls the trade-off between AUROC and FLOPs. A higher α prioritizes increasing AUROC, while a lower α prioritizes reducing FLOPs. Here, it is set to 0.9 as the impact of the AUROC remains more important while convergence is achieved. In both architectures, an output layer with a single neuron with the sigmoid function is used. ReLu is used as activation function, the models are trained using the Adam optimizer with a learning rate equal to 0.005, and the data are standardized beforehand. Subsequently, models with the highest AUROC are retained. The implementation of the proposed neural architecture search approaches was achieved using PyTorch and the Neural Network Intelligence toolkit [11]. Here, the search space and the search strategy are provided in configuration files, and the objective function is implemented and integrated to the code and used to calculate the performance of each architecture learned. The optimization is then performed by the toolkit.

4 Results

With the manually crafted experiment, $\text{MAN} - \text{MLP}_{\text{EUC}} = (l = 3, h_1 = 96, h_2 = 192, h_3 = 96)$ and $\text{MAN} - \text{MLP}_{\text{DTW}} = (l = 3, h_1 = 10, h_2 = 20, h_3 = 10)$ are retained as the optimal architectures with the EUC and the DTW similarity measures, respectively. With the NAS-based approach, the optimal architectures obtained with EUC and DTW are: $\text{NAS-MLP}_{\text{EUC}} = (l = 3, h_1 = 36, h_2 = 49, h_3 = 175)$ and $\text{NAS-MLP}_{\text{DTW}} = (l = 2, h_1 = 16, h_2 = 8)$, respectively. We notice that both approaches converged towards architectures with a high-dimensional projection of the features that helps with the separation. Table 1 presents the performance of the different models based on AUROC, true positive rate (TPR), false positive rate, FLOPs, and size of the models (number of parameters). The NAS approach has shown better performance in identifying optimal architectures for both EUC and DTW. $\text{NAS-MLP}_{\text{EUC}}$ achieved an AUROC of 0.995, with an FPR about 10% smaller than with $\text{MAN-MLP}_{\text{EUC}}$, the model is also about 70% lighter than the manually-learned one. Although the difference based on detection performance with DTW leans toward the manual model, it is minimal, and both the size and the FLOPs of $\text{NAS-MLP}_{\text{DTW}}$ are significantly smaller compared to $\text{MAN-MLP}_{\text{DTW}}$.

Visualizing and exploiting results of the anomaly detection helps to gain and communicate insights. *Postmortem* data, consisting of hundreds of RF pulses, are collected daily and saved to HDF5 files corresponding to station events. The ML-assisted analysis provides a comprehensive view of the stations and SRFCs affected within a rolling 24-h period, as illustrated in Figure 4 (Left). This is useful for continuously diagnosing short-term issues and addressing them promptly by the LLRF experts. On the Figure, we can read for example, that one cavity is affected from station A4 on the 25th of March 2024 at 13:33. This has generated two faulty pulses and these are identified as quenches. We are also tracking the fault labeled as "missing signal", which is triggered when one of the RF signals used in the GLR computation is not available, leading to erroneous GLR waveforms. These daily results are saved and rendered through the web and emails where more details are given, especially plots of the RF signals, to help the LLRF experts to gain a clearer understating of the anomalies. Long-term insights help capture patterns over time and space. For instance, identifying which SRFCs quench more frequently and determining when we experience more quenches. For the latter, as depicted in Figure 4 (Right), which has been obtained based on expert-corrected AI findings, a positive correlation is noticed between the operational energy and the number of quenches, as higher energies scheduled at the end of each half-year usually induce more quenches, in 2022, it is clearer in the second half of the year, where in November a total of 18 quenches has been recorded. Exceptions happen, for instance in April where the number of quenches is 21.

5 Discussion

The handcrafted experiment led to a directed but limited search over architectures that satisfy the provided structural constraints. In contrast, the ability of NAS to explore and evaluate a wide range of architectures explains its better performance. The current implementation is based on software running on offline data, a firmware implementation for an online deployment on FPGA has also been initiated. To meet the constraints for edge deployment while maintaining high detection performance, the NAS has shown to be an effective approach for lightweight model learning. The

TABLE 1 Evaluation results of the different MLPs.

Model	TPR	FPR	AUROC	FLOPs/Size
NAS – MLP_{EUC}	0.9861	0.0049	0.9950	11,367/10,847
MAN – MLP_{EUC}	0.9861	0.0055	0.9903	38,305/37,537
NAS – MLP_{DTW}	0.9583	0.0259	0.9662	241/193
MAN – MLP_{DTW}	0.9583	0.0257	0.9663	551/471

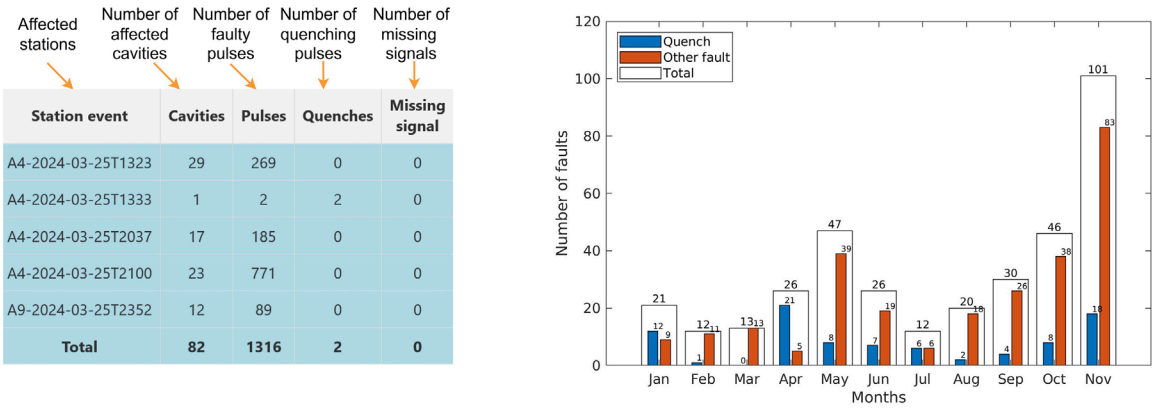

FIGURE 4
Leveraging AI output to gain insights across different time scales. (Left) Summary of SRFC anomalies from daily reports on 25 March 2024 where the affected stations, the number of faulty cavities and pulses, in addition to the type of anomalies are shown. (Right) Distribution of quenches and other SRFC anomalies from the yearly reports for 2022.

overall AUROC has also been maintained acceptable, and slightly improved with EUC, mainly as an impact of the number of false positives. A deeper analysis of a tolerable rate of false alarms is however needed to finalize the model selection. Additionally, the specifications of the targeted firmware need to be explored and eventually incorporated into the learning loop. More algorithms can also be explored to detect additional SRFC anomalies, with an emphasis on human-AI collaborative approaches to achieve adaptive and continual learning. Incorporating these new anomalies to the daily reports in Figure 4 would help the experts to quickly identify the eventual root cause or the affected subsystems. Moreover, this will help understand the other faults of the long-term analysis and find correlations.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

LB: Writing – original draft, Writing – review and editing. JB: Writing – review and editing. AE: Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded in the context of the R&D program of the European XFEL.

Acknowledgments

The authors acknowledge support from DESY (Hamburg, Germany), a member of the Helmholtz Association HGF. The authors thank Jan Horst Karl Timm, Christian Schmidt and Nicholas Walker for their input.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

1. Einstein-Curtis J, Drees K, Edelen J, Kilpatrick M, Laster J, O'Rourke R, et al. Classification and prediction of superconducting magnet quenches. In: *19th international conference on accelerator and large experimental Physics control systems* (2024). p. 856–9.

2. Rahman MM, Carpenter A, Iftekharuddin K, Tennant C. Accelerating cavity fault prediction using deep learning at jefferson laboratory. *Machine Learn Sci Technology* (2024) 5:035078. doi:10.1088/2632-2153/ad7ad6

3. Branlard J, Ayvazyan V, Hensler O, Schmidt C, Schlarb H. Superconducting cavity quench detection and prevention for the European XFEL. In: *14th international conference on accelerator and large experimental Physics control systems* (2013). p. 1239–41.

4. Eichler A, Branlard J, Timm J. Anomaly detection at the european X-ray Free Electron Laser using a parity-space-based method. *Phys Rev Acc Beams* (2023) 26:012801. doi:10.1103/PhysRevAccelBeams.26.012801

5. Nawaz A, Pfeiffer S, Lichtenberg G, Rostalski P. Anomaly detection for the European XFEL using a nonlinear parity space method. In: *10th IFAC symposium on fault detection, supervision and safety for technical processes* (2018). p. 1379–86.

6. Boukela L, Eichler A, Branlard J, Jomhari NZ. A two-stage machine learning-aided approach for quench identification at the European XFEL. In: *12th IFAC symposium on fault detection, supervision and safety for technical processes* (2024). p. 402–7.

7. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *nature* (1986) 323:533–6. doi:10.1038/323533a0

8. Schilcher T. *Vector sum control of pulsed accelerating fields in Lorentz force detuned superconducting cavities*. Hamburg, Germany: University of Hamburg (1998).

9. Kaufman L, Rousseeuw PJ. *Finding groups in data: an introduction to cluster analysis*. New York, USA: John Wiley and Sons, Inc (1990).

10. Elsken T, Metzen JH, Hutter F. Neural architecture search: a survey. *J Machine Learn Res* (2019) 20:1–21. doi:10.5555/3322706.3361996

11. Microsoft. Neural network intelligence (NNI) (2024). Available online at: https://github.com/microsoft/nni (Accessed January, 2024).