

Coincident learning for beam-based rf station fault identification using phase information at the SLAC linac coherent light source

Jia Liang^{1,*}, William Colucho², Franz-Josef Decker², Ryan Humble³, Ben Morris², Finn H. O'Shea², David A. Steele², Zhe Zhang², Eric Darve^{1,‡} and Daniel Ratner^{2,†}

¹*Institute for Computational and Mathematical Engineering, Stanford University, Stanford, California 94305, USA*

²*SLAC National Laboratory, Menlo Park, California 94025, USA*

³*Cerebras Systems, Sunnyvale, California 94085, USA*



(Received 19 May 2025; accepted 30 October 2025; published 3 December 2025)

Anomalies in radio-frequency (rf) stations can result in unplanned downtime and performance degradation in linear accelerators such as SLAC's Linac Coherent Light Source (LCLS). Detecting these anomalies is challenging due to the complexity of accelerator systems, high data volume, and scarcity of labeled fault data. Prior work identified faults using beam-based detection, combining rf amplitude and beam position monitor data. Due to the simplicity of the rf amplitude data, classical methods are sufficient to identify faults, but the recall is constrained by the low-frequency and asynchronous characteristics of the data. In this work, we leverage high-frequency, time-synchronous rf phase data to enhance anomaly detection in the LCLS accelerator. Due to the complexity of phase data, classical methods fail, and we instead train deep neural networks within the Coincident Anomaly Detection (CoAD) framework. We find that applying CoAD to phase data detects nearly 3 times as many anomalies as when applied to amplitude data, while achieving broader coverage across rf stations. Furthermore, the rich structure of phase data enables us to cluster anomalies into distinct physical categories. Through the integration of auxiliary system status bits, we link clusters to specific fault signatures, providing additional granularity for uncovering the root cause of faults. We also investigate interpretability via Shapley values, confirming that the learned models focus on the most informative regions of the data and providing insight for cases where the model makes mistakes. This work demonstrates that phase-based anomaly detection for rf stations improves both diagnostic coverage and root cause analysis in accelerator systems and that deep neural networks are essential for effective analysis.

DOI: [10.1103/zmmr-ry9h](https://doi.org/10.1103/zmmr-ry9h)

I. INTRODUCTION

Identifying and resolving anomalies in particle accelerators can improve both uptime and performance for users. The Linac Coherent Light Source (LCLS) experiences unplanned downtime and unexpected behavior, reducing beam stability and introducing noise sources that can interfere with user data analysis. Among the various failure modes of LCLS, our primary concern is pinpointing anomalous behavior in the 82 radio-frequency (rf) stations that power the x-ray laser's accelerator. Anomalies in

the rf stations are an interesting source of study for two reasons [1]: first, they are a common failure mode, with several drops in rf station performance per hour during beam operation. Second, they have a significant impact on performance because faults directly affect the beam energy, causing fluctuations in the pulse energy and photon energy of the free-electron laser delivered to users.

The increasing complexity of particle accelerators has made it difficult for humans to monitor and manage faults effectively. For example, LCLS's control system manages 200,000 process variables, highlighting the need for data science techniques to enhance the automation of accelerator operations. To alleviate the burden on human operators, our previous work [1] proposed a beam-based method to identify changes in accelerator status by observing shot-to-shot data from the beam position monitoring system. By comparing these beam-based anomalies with data from individual rf stations, we can determine which rf station was responsible for the change in beam behavior. The classical statistical method from [1] can be effective, but requires significant effort and manual data inspection to

*Contact author: jialiang@slac.stanford.edu

†Contact author: dratner@slac.stanford.edu

‡Also at Mechanical Engineering, Stanford University, Stanford, California 94305, USA.

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

choose the right hyperparameters. Shifts in data distribution overtime can make the selected hyperparameters—and, consequently, the algorithm—ineffective. More critically, the classical approach has only been applied to the low-rate amplitude data stream. The full-rate phase data stream are expected to include far more information about anomalies, but is too complex to be analyzed by classical methods. Consequently, phase data were not used in [1]. To illustrate those points, in Appendix A, we present several subtle examples in which Coincident Anomaly Detection (CoAD) accurately identifies anomalous and nonanomalous instances—cases that would be difficult to detect using the classical anomaly detection method.

Compared to amplitude data, we find that working with phase data offers several advantages. First, the phase data captures new information, enabling detection of anomalies that amplitude-based detection misses. Second, the time-synchronous phase data provide critical insights, allowing us to distinguish whether a change in the accelerator status occurs simultaneously with an rf station change or in response to a previous issue. Third, the rich information contained in the phase data facilitates clustering of anomalies into distinct categories, each with unique signatures. This categorization can aid in the expert diagnosis of the root cause within the rf stations and expedite recovery.

In our previous work [2], we found that CoAD is an effective approach to detect rf faults. CoAD is designed to identify anomalies in multimodal systems by leveraging coincident behavior across two data slices, e.g., an individual subsystem’s performance and an overall system output “quality.” Any subsystem issue should manifest as coincident anomalies in both the subsystem and the larger system output, and this behavior can be exploited to train deep learning anomaly detection algorithms [2]. In the context of the rf particle accelerator task, for example, one data stream comes from a radio-frequency (rf) station subsystem (s), while another data stream originates from beam position monitors (BPMs), which provide data on the final electron beam energy and quality (q). Our previous work demonstrated the effectiveness of CoAD in real-world datasets.

The previous studies on rf stations produced reasonable results using time-asynchronous amplitude data, but ignored the richer information from time-synchronous phase data, which, due to its complexity, could not be analyzed by the statistical method from [1]. In this paper, we extend the earlier work by training deep neural networks to analyze the phase data.

This paper makes three primary contributions. First, we present the first automated framework for analyzing rf phase data, enabled by deep neural networks (DNNs), which eliminates the need for manual inspection and facilitates scalable anomaly detection in particle accelerators. Second, we apply our method to operational data from the LCLS and demonstrate a substantial improvement in detecting rf anomalies, significantly outperforming existing baseline

approaches. Third, we show that phase signatures exhibit distinct clustering patterns that align with known fault types, providing additional diagnostic insights and enhancing the interpretability of the anomalies by revealing potential root causes. The rest of the paper is divided into the following sections: Sec. II introduces the relevant background and related work; Sec. III introduces rf anomalies and compares phase data with amplitude data. Section IV presents the proposed methodology; Sec. V discusses the experimental results; and Sec. VI concludes the paper and outlines future work.

II. RELATED WORK

Machine learning (ML) techniques have been extensively applied to fault detection in particle accelerators. For instance, [3] introduced an unsupervised long-short term memory (LSTM) autoencoder-based anomaly detection framework for the CEBAF accelerator’s orbit lock system, which achieved high fault detection rates by modeling normal operational patterns to detect deviations. Similarly, [4] developed a Smart Alarm system for the CEBAF injector beamline based on a neural network inverse model that predicts machine settings from sensor readings. This system detected 83% more anomalous conditions than the existing rule-based system and achieved 94.6% accuracy in identifying the root cause of faults through discrepancies between predicted and observed settings.

Recent works [5–15] in this domain tend to focus on *early fault detection* and *fault prevention*, aiming to issue alerts *prior* to the actual onset of faults to enhance machine protection and reduce operational downtime. For example, [11] proposed a deep learning architecture combining LSTM and convolutional neural network (CNN) to predict slowly developing faults in accelerating cavities at CEBAF, achieving 80% accuracy up to 1 s before fault occurrence. Their model leverages multivariate rf signal data and enhances predictive precision through a consecutive window criterion and confidence thresholding. In a similar vein, [5] demonstrated the utility of temporal deep learning models such as LSTM and temporal convolutional networks for identifying prefault signatures in rf control systems.

Most aforementioned approaches [5,6,8–13] rely on supervised learning, which necessitates labeled datasets. A subset of recent works—such as those proposed in [3,4,7,14,15]—explores unsupervised strategies, but these typically assume that the training data is entirely normal. For example, [3] had to manually exclude faulty data during training to uphold this assumption. However, a key challenge in the fault detection for rf stations is the lack of labeled fault data. In such settings, unsupervised anomaly detection methods are a natural choice, yet their effectiveness is often undermined when the training data is contaminated with unidentified anomalies or noise—a common scenario in particle accelerator environments. As a result, their performance can degrade significantly

in practice. In contrast, the method presented in this work requires neither labeled data nor a purely normal training set. Our approach is robust to significant anomaly contamination in the training dataset.

Another emerging direction in this area is the integration of advanced ML techniques to enhance *performance*, *adaptability in low-data regimes*, and *model robustness*. Regarding performance optimization, [14] employed neural architecture search and hyperparameter tuning, resulting in an 8% increase in the true positive rate for errant beam detection. For adaptability in data-scarce scenarios, [13] introduced a bidirectional discriminative prototype network for few-shot fault diagnosis in particle accelerator power systems. Their approach trains solely on normal operational data and excels in multiclass fault diagnosis, outperforming conventional deep learning models under limited fault sample conditions. For model robustness, [6] proposed an uncertainty-aware Siamese neural network to predict errant beam pulses at the Oak Ridge Spallation Neutron Source, achieving low false positive rates while also providing calibrated uncertainty estimates. Similarly, [11] employed Monte Carlo dropout at inference time to quantify predictive uncertainty and improve model reliability. In our work, we leverage a CoAD framework to reduce uncertainty, requiring agreement across multiple data streams before confirming an anomaly.

Methodologically, the works of [6] and [9] bear similarities to our proposed CoAD framework. Both employ Siamese networks, which, like CoAD, utilize dual input streams. However, key differences distinguish our approach. Siamese networks require labeled input pairs—typically “normal-normal” or “normal-abnormal”—whereas CoAD operates without any labeled data. Furthermore, Siamese networks typically use inputs from the same modality (e.g., the same sensor type), while CoAD is designed for multimodal learning, requiring inputs from heterogeneous sources, such as system-level beam data and subsystem-specific diagnostic signals. Moreover, CoAD addresses a unique and critical challenge. In complex systems like particle accelerators, diagnostic signals are often noisy and contain numerous anomalies, most of which are benign or unrelated to true faults. CoAD is specifically designed to disentangle these benign anomalies from fault-indicative anomalies. Unlike conventional anomaly detection methods, which merely identify deviations from normal behavior, CoAD leverages coincident learning to differentiate anomalies that genuinely signal faults. This capability is essential for fault diagnosis in such environments, where simply detecting anomalies is insufficient. Additionally, at SLAC, recorded anomalies are extremely rare. Directly training on these rare events introduces further complications. Beyond their scarcity, most recorded anomalies correspond to “hard faults” that are easily recognizable by human operators. Supervised training on such data risks biasing the model toward detecting only these overt failures, potentially

overlooking subtle anomalies that escape operator attention. A limitation, however, is that CoAD can only detect anomaly patterns present in the training data, making it more effective for recurring faults than for one-off, outlier events.

III. RF ANOMALIES

The performance of LCLS is evaluated by the quality of the x-ray beam it delivers, measured through factors including pulse energy, photon energy stability, and bandwidth. As in our previous work [1], an anomaly is defined as any unforeseen change in the accelerator condition that ranges from diminishing the quality of the beam to a complete loss of beam. Among the various failure modes, we focus on identifying anomalies in the 82 rf stations because they power the x-ray laser’s accelerator, directly impact beam energy, and represent a significant source of failures.

A. Amplitude data

The existing method leverages two sources of data: rf station diagnostics, which provide insight into fault origins, and beam-based diagnostics, which capture the pulse-by-pulse performance of the accelerator.

The previous work used two rf station diagnostics: the real-valued amplitude signals (AMPL) or the summary amplitude mean out of tolerance status bit (AMM), which indicates whether the rf station’s amplitude is outside a predefined range. Both the AMM and AMPL data are recorded asynchronously with the beam data at a rate of about 0.2 Hz and can lag behind the beam data by several seconds.

The beam-based diagnostic data originates from stripline beam position monitors (BPMs) [16] located in dispersive regions, which are sensitive to fluctuations in beam energy. The BPM data are recorded synchronously at the full LCLS operational rate of 120 Hz. The accelerator contains 175 BPMs, each measuring the beam’s transverse position (X , Y), and charge (TMIT). Among these 175 BPMs, our work focuses on four specific BPMs in dispersive regions: BPMS:LTUH:250, BPMS:LTUH:450, BPMS:DMPH:502, and BPMS:DMPH:693. For each selected BPM, we use its charge measurement along with either the X or Y position (depending on the dispersive direction), as shown in Fig. 1. We select these four BPMs because they are located in regions with nonzero horizontal or vertical dispersion, rendering them particularly sensitive to beam energy variations. Moreover, their placement downstream of all rf stations enables them to capture cumulative effects from upstream components. Previous work in [1,2] cross-reference BPM data with rf AMM/AMPL data to detect rf anomalies.

B. Phase data

In this work, we aim to leverage phase data from the rf station subsystem(s) to detect rf anomalies. To ensure stable acceleration, the rf oscillations must align precisely with the periodic arrival of the particle beam in the cavity.

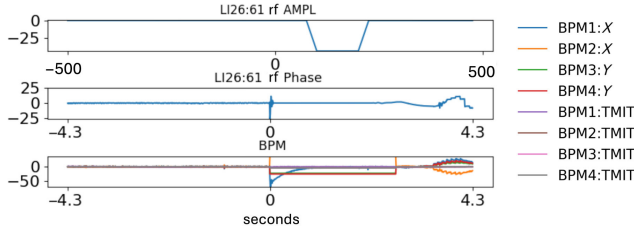


FIG. 1. Comparison of rf amplitude and rf phase diagnostic data during an rf fault in station LI26:61. The fault impacted downstream beam quality, as evidenced by the beam position monitor (BPM) signals. BPM and rf phase data are synchronous, and the data reflects correlated behavior. The rf amplitude diagnostic is slow, reporting only a step function change, and asynchronous with a multisecond lag relative to the phase and BPM data. Specifically, BPM1 is BPMS:LTUH:250, BPM2 is BPMS:LTUH:450, BPM3 is BPMS:DMPH:502, and BPM4 is BPMS:DMPH:693.

The particle’s arrival time relative to the rf cycle is referred to as the rf phase. Compared to amplitude data, phase data offers several advantages. First, the phase data are recorded at the full LCLS operation rate of 120 Hz, which matches the sampling rate of the BPM data. Moreover, the phase data are synchronized with the BPM data. A comparison between the phase and amplitude data is summarized in Table I, and an example of rf phase and rf amplitude can be seen in Fig. 1. The figure illustrates a clear relationship between variations in phase data and corresponding changes in BPM behavior. Notably, at the 0-s mark, a distinct jump in the phase data aligns with a corresponding jump in the BPM. Toward the end of the observation window, the curved behavior of the rf phase also resembles the curved patterns observed in the BPM. In contrast, jumps in the AMPL data can occur more than 5 s apart from those in the BPM.

IV. METHOD

Our method consists of two main steps: “candidate” generation and CoAD. The candidate generation stage reduces the data rate and enhances the anomalous fraction of the dataset fed to CoAD. Candidate generation itself involves two substeps: we first use beam data to identify time windows with potentially anomalous behavior, and then use rf station diagnostic data to identify potentially

TABLE I. Comparison between rf phase diagnostic data and rf amplitude diagnostic data.

	Prior works [1,2]	Current
Data streams	Amplitude (s) BPM (q)	Phase (s) BPM (q)
Sampling rate	s (~ 0.2 Hz) q (120 Hz)	s (120 Hz) q (120 Hz)
Synchronicity	s lags behind q up to 5 s	s and q are synchronous

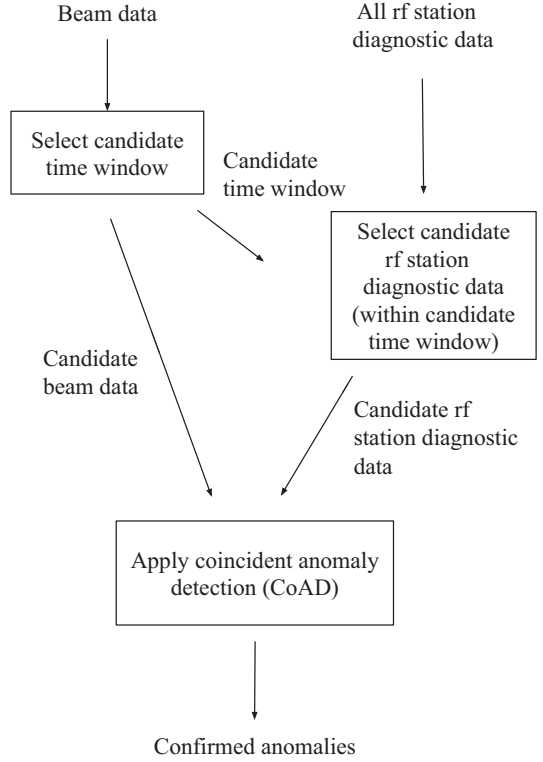


FIG. 2. Anomaly detection workflow: beam data are used to identify candidate time windows with potential anomalies, and rf diagnostic data highlights suspect rf stations. CoAD is then applied to detect anomalous behavior within the candidate windows.

anomalous rf stations within each candidate window. For each such window, we pair the corresponding beam data with the diagnostic data from the selected rf stations to form anomaly candidates. CoAD is then applied to these selected candidates to identify true anomalies. This process is illustrated in Fig. 2.

A. Anomaly candidate generation

The candidate generation step uses beam data and rf stations’ diagnostic data to generate anomaly candidates. The beam data are the multivariate time-series BPM data $x_{s,t}$ for signal s at time t . The anomaly score is defined as $a_{s,t} = \text{MAD}(x_{s,t-2l}, \dots, x_{s,t})$, where l is the lagging rolling window size for computing the median and median-absolute deviation (MAD). Then, we aggregate the per-signal anomaly scores into a single anomaly score a_t using the geometric mean:

$$a_t = \left(\prod_{s=1}^n a_{s,t} \right)^{\frac{1}{n}},$$

where n is the number of signals. We further aggregate across k consecutive pulses to get the final score

$$a_{\text{AGG},t} = \left(\prod_{i=0}^{k-1} a_{t-i} \right)^{1/k}$$

to boost the score for sustained beam anomalies. The parameter k controls the emphasis on sustained anomalies; higher values of k prioritize longer-duration anomalies.

We identify a candidate C_i centered at time t_i if the aggregated anomaly score a_{AGG,t_i} meets or exceeds a predefined threshold τ , i.e., $a_{\text{AGG},t_i} \geq \tau$. However, since the aggregated score $a_{\text{AGG},t}$ is compared with the threshold τ , the time point t_i satisfying $a_{\text{AGG},t_i} \geq \tau$ may not correspond to the earliest occurrence of the anomaly. To address this, we employ a simple heuristic to determine the earliest time point t'_i within the sequence where the anomaly first occurs. Specifically, we identify t'_i as the first point within the range $(t_i - k, t_i)$ whose anomaly score, $a_{t'}$, deviates by at least 1.25 standard deviations from the mean of the anomaly scores of all points in that range. We then use t'_i as the center of the anomaly candidate C_i .

Once the center t'_i is determined, the full data window is defined as the range $(t'_i - w, t'_i + w)$, where w is the window size. In this work, the window size is set to 533 samples, corresponding to roughly 4.3 s. This duration is chosen because beam faults can persist for several seconds, and recovery from an initial fault may also take a few seconds. Extending the window further does not provide significant additional benefits. The candidate C_i consists of two components. The first component is the beam data, which includes eight channels of BPM data as described in Sec. III A. The second component is the rf diagnostic data, specifically the phase data. In our approach, we use the phase data from the rf station(s) exhibiting the most anomalous behavior around the center of the time window to identify the most likely root cause of the anomalies. In this work, the most anomalous rf stations are identified by measuring how much their phase values deviate from zero within a short time window preceding the beam trigger. The algorithm selects the five stations with the largest deviations, retains only those exceeding a predefined threshold, and pairs them with beam data for further analysis. If none exceeds the threshold, the station with the highest deviation is chosen. Additional details on the identification procedure are provided in Appendix D. Formally, let us denote the rf diagnostic data as s , describing the subsystem, and the BPM data as q , representing the overall quality of the system. Consequently, each candidate C_i is defined as $C_i = \{s_i, q_i\}$. Please refer to Appendix F for the list of hyperparameters used in the candidate generation process, along with their values and the rationale for each choice.

B. Coincident anomaly detection

Coincident Anomaly Detection (CoAD) [2] is an unsupervised anomaly detection approach designed for datasets with paired inputs $D = \{(s, q)\}$, where s and q are two

streams of data. In the context of rf station fault detection, s represents diagnostic data from the rf station subsystem, and q represents a system-level quality indicator, such as measurements from beam position monitors. CoAD addresses tasks where fluctuations in s and q are uncorrelated during normal operating conditions, but both data streams are expected to change simultaneously during an anomalous event. By leveraging this coincidence principle, CoAD overcomes challenges faced by existing anomaly detection techniques, such as requiring labeled datasets or datasets known to contain only normal behavior.

CoAD is particularly useful in multimodal systems, where data comes from two distinct but related sources—such as different sensors monitoring the same system or different stages of a sequential process. Furthermore, the algorithm remains applicable even with a single measurement, provided that two inputs, s and q , can be derived by partitioning the data.

From a technical perspective, coincident learning utilizes two models, $A_{\theta_s}(s)$ and $A_{\theta_q}(q)$, parameterized by θ_s and θ_q , respectively. Each model outputs a scalar value, p_s and p_q , where $p_s, p_q \in [0, 1]$, representing the confidence that the corresponding input is anomalous. The marginal datasets are $D_s = \{s\}$ and $D_q = \{q\}$. Both s and q are assumed to be functions of an unknown variable x , denoted as $s(x)$ and $q(x)$. A schematic representation of the method is shown in Fig. 3.

The goal of learning is to optimize A_{θ_s} and A_{θ_q} to maximize the alignment between the outputs p_s and p_q . Previously proposed unsupervised metrics include covariance, correlation, and $\hat{F}_\beta$. Specifically, $\hat{F}_\beta$ is defined as follows:

$$\hat{F}_\beta = (1 + \beta^2) \frac{\mu_{sq} - \mu_s \mu_q}{\mu_{sq} + \alpha \beta^2} \frac{1 - \mu_{sq}}{(1 - \mu_s)(1 - \mu_q)},$$

where α is the anomaly fraction in the data, $\mu_{sq} = E[A_{\theta_s}(s)A_{\theta_q}(q)]$, $\mu_s = E[A_{\theta_s}(s)]$, and $\mu_q = E[A_{\theta_q}(q)]$. An estimate of α is sufficient. It has been shown that $\hat{F}_\beta$ is a

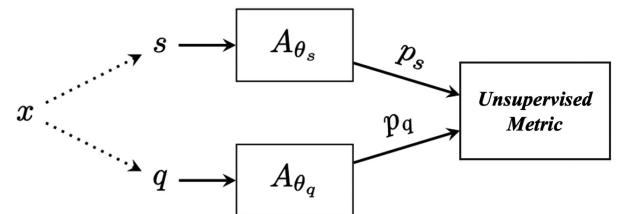


FIG. 3. Generic CoAD schematic with an unknown state variable x influencing measured data inputs s and q fed to corresponding models, A_{θ_s} and A_{θ_q} . The models are trained to optimize an unsupervised objective metric such as correlation, covariance, or $\hat{F}_\beta$ such that the models maximally separate x into normal and anomalous behavior [2].

lower bound for F_β in [2]. In this work, we use the F_β objective with the choice of $\beta = \infty$, which is equivalent to the covariance objective. This choice of β emphasizes recall and improves training stability. In general, CoAD holds a distinct advantage over other methods because its loss function is fully unsupervised, enabling direct hyperparameter optimization against the validation dataset. In practice, we have observed that CoAD is largely insensitive to hyperparameter choices. Additionally, the hyperparameters mentioned in candidate generation step (Sec. IV A) primarily affect the fraction of anomalies and do not have a strong influence on results.

V. EXPERIMENTS

A. rf phase vs rf amplitude diagnostic data

We evaluate the performance of our two-stage rf station anomaly detection method, shown in Fig. 2, using both rf phase and amplitude data over the period from January 18, 2024, to January 30, 2024. To simplify data processing, we focus exclusively on operational periods during which the beam was delivered to experiments under healthy conditions [1]. Additionally, we exclude time periods when the beam-synchronous acquisition data-recording service exhibited anomalous performance, identified using the Bayesian online changepoint detection algorithm [17]. We follow the same data filtering protocol as [1]. After applying these criteria, we analyze approximately 230 h of data within the specified period.

Within the selected period, the proposed method generates 5383 anomaly candidates using rf amplitude data and 5364 candidates using rf phase data. After excluding system-level faults—faults in which all rf stations exhibit abnormal behavior, preventing attribution to specific stations—the proposed method using rf amplitude diagnostic data identifies 209 anomalies, 185 of which are confirmed by manual inspection by experts as true anomalies, resulting in a precision of 88.52%. Similarly, the same method using rf phase diagnostic data predicts 548 anomalies, of which 482 are true anomalies, yielding a precision of 87.95%. These results are presented in Table II. Additionally, after examining 300 negative examples predicted by both methods, we observed 3 and 4 false negatives for rf amplitude data and rf phase data, respectively, implying that the false negative rate for each is approximately 1%. Both methods predict approximately 5000 examples as negatives, with around 50 events missed by each. This results in a recall¹ of 87% for the phase-based method and 78% for the amplitude-based method. Next, we examined the overlap and distinctiveness of anomalies identified by both methods. We consider rf anomalies identified using amplitude diagnostic data to be equivalent

TABLE II. Performance evaluation of CoAD in detecting rf anomalies using phase vs amplitude diagnostic data: a case study from January 18 to January 30, 2024. Note that 62 anomalies can be identified by both rf amplitude and rf phase diagnostic data and are accounted for in both lines.

	No. of anomalies	Precision (%)	Recall (%)
Phase-based method	482	87.95	87
Amplitude-based method	185	88.52	78

to those identified using phase diagnostic data if the following conditions are satisfied: (1) the anomalies share the same time windows for the BPM data, and (2) the faults are associated with the same rf station. Based on these criteria, 123 anomalies were unique to rf amplitude diagnostic data, while 420 anomalies were unique to rf phase diagnostic data, with 62 anomalies identified by both rf amplitude and rf phase diagnostic data. Consequently, by integrating rf phase diagnostic data with amplitude diagnostic data, anomaly detection improves by more than threefold, identifying 605 anomalies using both data streams compared to just 185 detected using amplitude data alone.

A detailed investigation of the distribution of anomalies across rf stations further underscores the benefits of incorporating rf phase diagnostic data. As illustrated in Fig. 4, anomalies identified exclusively by rf amplitude diagnostic data are predominantly concentrated in specific rf stations, such as LI24:51, LI25:11, LI26:71, and LI27:11–81. A closer analysis of LI27:11–81 reveals that phase data are absent across all LI27 rf stations, explaining why anomalies in these stations can only be detected using rf amplitude diagnostic data. In contrast, anomalies detected solely through rf phase diagnostic data are distributed across a broader range of rf stations. Notably, certain rf station faults, such as those in 20:61–81, can only be identified using rf phase data. These stations are particularly critical for anomaly detection, as they are located near the injector, where there are fewer spare rf stations to compensate for faulty behavior. This finding highlights that incorporating rf phase diagnostic data not only increases the overall detection capability but also enhances the diversity of detected anomalies in terms of rf station coverage.

An additional advantage of utilizing rf phase diagnostic data is its ability to identify some types of jitter, e.g., the appearance of frequent, small spikes in phase, which would be undetectable using amplitude data alone. Further details on repeated spike faults are provided in Appendix B.

B. Comparison of different anomaly detection methods

For both the rf amplitude diagnostic data and rf phase diagnostic data, we employ CoAD as our *de facto* anomaly

¹It is important to note that recall is calculated based on the anomaly candidate dataset generated in Sec. IV A.

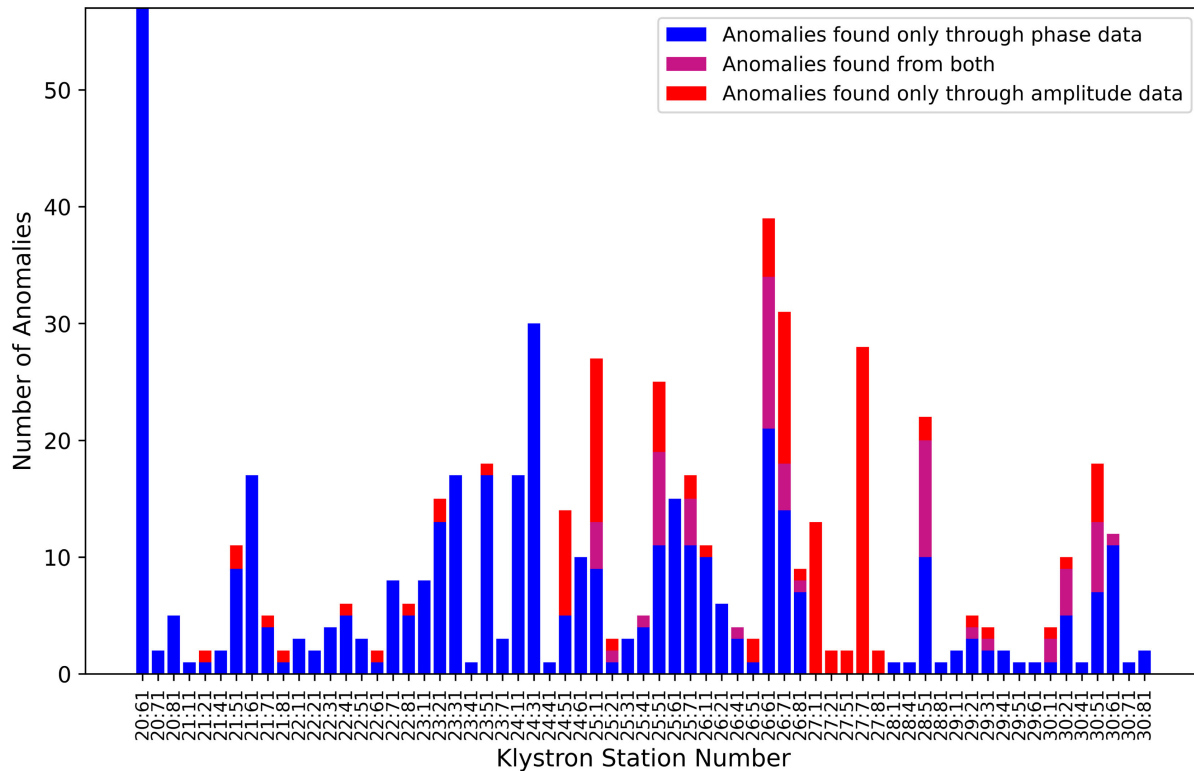


FIG. 4. Distribution of the anomalies in rf stations found using CoAD. The colors correspond to the data source: red for anomalies found only from amplitude data, blue for anomalies found only from phase data, and magenta for anomalies found in both data streams. Sector 27 has no phase anomalies due to a software fault preventing logging of phase data during this time period. Note that the critical sector 20 injector station anomalies are only found from phase data.

detection method. In this section, we evaluate the effectiveness of CoAD in comparison to other anomaly detection methods. Training details for CoAD are provided in Appendix E. Specifically, we focus on rf phase data for this evaluation. The following list details the methods that are compared: (i) CCA-1D: canonical correlation analysis (CCA) is a statistical method employed to identify and quantify the correlation between two sets of variables. We utilize the implementation provided by the library [18]. To apply this method, it is necessary to convert our 2D data into a 1D format. We include the CCA method because it can be regarded as the linear counterpart of CoAD. (ii) CCA-2D: similar to the above, except this method accommodates 2D data. We employ the implementation from the library [19]. (iii) IForest [20]: an unsupervised machine learning algorithm that identifies anomalies by recursively partitioning the data using randomly selected features and split values, effectively isolating outliers due to their shorter paths in the tree structure. (iv) OCSVM [21]: an unsupervised learning algorithm that constructs a decision function for outlier detection by estimating a high-dimensional boundary around the data's normal class, effectively separating outliers based on their deviation from this boundary. (v) DGHL [22]: employs a top-down convolutional network to map multivariate time-series

windows to a hierarchical latent space, using alternating backpropagation and short-run Markov chain Monte Carlo for training by maximizing observed likelihood, thus enabling efficient anomaly detection and robust handling of missing data. (vi) OmniAnomaly [23]: a robust multivariate time series anomaly detection method using a stochastic recurrent neural network that leverages techniques like stochastic variable connection and planar normalizing flow to learn and reconstruct normal patterns.

AUCPR and best F_1 score² are commonly used metrics to evaluate the performance of anomaly detection algorithms. In Table III, we report both metrics for all previously discussed anomaly detection methods. These metrics do not require selecting a decision threshold—they summarize performance across all possible thresholds. However, in deployment scenarios, a specific threshold must be chosen to classify inputs as normal or anomalous. In Table IV, we examine how sensitive each method is to the choice of threshold and show that CoAD maintains stable performance across a wide range of thresholds. All experiments use 80% of the data for training and 20% for

²It is important to note that both AUCPR and the best F_1 score are calculated based on the anomaly candidate dataset generated in Sec. IVA.

TABLE III. Comparison of different anomaly detection methods on phase-based diagnostic data. CoAD shows superior performance under both AUCPR and F_1 . Bold values indicate the best performance among all compared methods within each column (i.e., the highest score for that metric).

		AUCPR	Best F_1
Traditional	CCA-1D	0.32	0.4
	CCA-2D	0.32	0.44
	IForest	0.44	0.41
	OCSVM	0.54	0.57
Deep learning	DGHL	0.85	0.75
	OmniAnomaly	0.73	0.71
	CoAD	0.91	0.89

testing. The split is performed temporally to reflect the online scenario: all training data precede January 28, 2024 14:07:37, and all testing data follow this timestamp. The data used in this experiment can be accessed at [24].

Our data consist of two input streams: rf diagnostic data and BPM data. Methods like CCA-1D, CCA-2D, and CoAD naturally accommodate both input streams. However, methods such as OmniAnomaly, DGHL, IForest, and OCSVM are designed to process only a single stream of data by default. To benchmark against methods not designed for multiple inputs, we attempted two strategies: Score & Stack (training and scoring on each input independently and using the Pareto frontier of all threshold pairs) and Stack & Score (training and scoring on the stacked input) [2]. For these methods, we ran both strategies and reported the score of the better-performing approach. The performance of all the methods using AUCPR and best F_1 metrics is shown in Table III. The table indicates that OCSVM achieves the highest performance among all traditional methods. OmniAnomaly outperforms all traditional methods, while DGHL ranks as the second-best method overall. Notably, CoAD demonstrates the best performance across all methods, achieving the highest AUCPR and best F_1 scores.

In real-world applications, setting a threshold is essential to convert anomaly scores into anomaly predictions. Therefore, a fair evaluation also requires the calculation of the F_1 score using various thresholding methods. To evaluate the performance of different anomaly detection methods using F_1 score under various thresholding strategies, we employ the following thresholding methods: Filter [25], Meta [26], and MixMod [27]. All of these thresholding methods are implemented in the `Pythresh` library, and the library’s comprehensive benchmark across various datasets and methods demonstrates that these chosen thresholding methods achieve superior performance. Since the F_1 score with thresholding methods will be worse than the best F_1 score, we calculated the F_1 score with various thresholding methods only for the deep learning methods. The results are shown in Table IV.

TABLE IV. Comparison of all the deep-learning-based methods using F_1 score with three different thresholding methods: Meta, MixMod, and Filter. In all cases, CoAD produces the best scores. Bold values indicate the best performance among all compared methods within each column (i.e., the highest score for that metric).

	Meta	MixMod	Filter
CoAD	0.83	0.79	0.87
OmniAnomaly	0.39	0.34	0.34
DGHL	0.44	0.54	0.33

From the table, it is evident that the CoAD method achieves good results with all three thresholding methods. Moreover, its F_1 scores are only slightly lower than the best F_1 score. In contrast, the F_1 scores for OmniAnomaly and DGHL with various thresholding methods are significantly worse than their best F_1 scores. This result indicates that CoAD is more robust at separating the anomalous and normal samples, thereby increasing the probability of success during deployment. Furthermore, because CoAD provides an unsupervised estimate of precision and recall, the CoAD framework can also be used to select a threshold for deployment even in the absence of labels.

C. Clustering and root-cause analysis

A key advantage of rf phase data is the high information content of the synchronous, beam-rate data compared to the slow, asynchronous amplitude diagnostic. The complexity of the rf phase data could reveal signatures of the underlying root cause, i.e., the ability not just to identify a specific problematic rf station but also *why* that station failed. Providing this information to operators would improve both the interpretability of warnings and the speed of diagnosis and recovery from faults.

To investigate the potential for root-cause analysis from rf phase diagnostic data, we undertake the following steps: we begin by collecting the rf phase signals labeled by CoAD as anomalous from the following time periods: February 7–23, 2022; March 5–17, 2022; April 1–16, 2022; November 9–18, 2023; November 21–December 1, 2023; and January 18–30, 2024. We then apply UMAP [28] to reduce the dimensionality of the raw signals to two dimensions and employ HDBSCAN [29] to cluster the 2D embeddings. The clusters are displayed in Fig. 5.

More importantly, in the same figure, we also show the relationship between the clusters and their association with different “status bits.” The monitoring system for each rf station records not only amplitude and phase information but also various status bits, including Digital Status 1 (DSTA1), Digital Status 2 (DSTA2), Status (STAT), Hardware Descriptor (HDSC), Hardware Status (HSTA), and PIOP Status (SWRD). Each status bit further contains substatus indicators, which are intended to reveal the

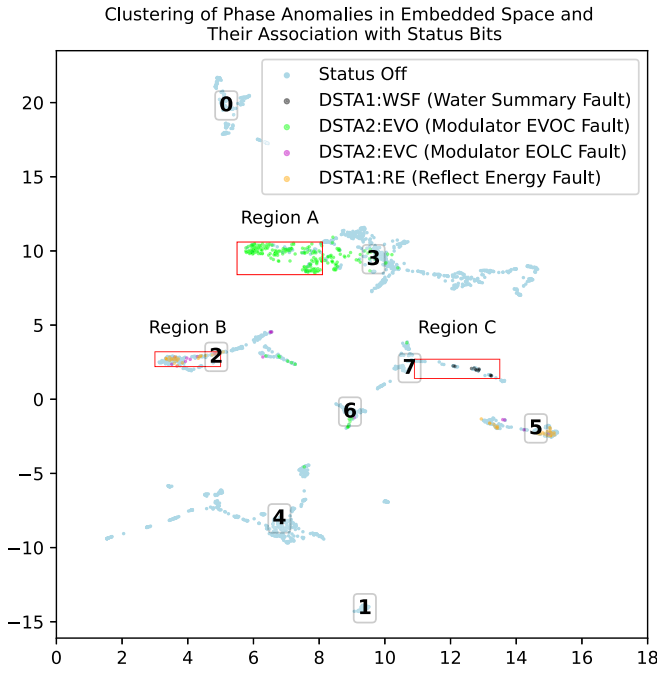


FIG. 5. Clustering of the anomalous rf phase signals and their association with status bits. Example phase and BPM data selected from regions A, B, C are shown in Fig. 6, with more details shown in Fig. 11.

underlying cause of a hardware problem. For example, DSTA1:RE signifies the Digital Status Reflected Energy fault, while DSTA2:EVO denotes a Modulator EVOC Fault. While highly specific, these status bits are known to have poor precision and recall, and thus cannot be used as an anomaly detection system by themselves.

To better understand fault patterns, we analyzed the activation and deactivation of these status bits during rf faults. While many status bits exhibited interesting trends, our discussion primarily focuses on the following: DSTA1:RE, DSTA1:WSF, DSTA2:EVC, and DSTA2:EVO. The distribution of these status bits across clusters is also illustrated in Fig. 5. The same figure also highlights several regions with red boxes, and their enlarged views are shown in Fig. 11 in Appendix C. For detailed definitions and the associated root causes of these status bits, see Appendix G. (i) *DSTA2:EVO (Modulator EVOC Fault—Electrode Voltage Overcurrent)*: These faults predominantly appear in the left part of cluster 3. Notably, part (a) of Fig. 6 reveals a distinct signature of rf phase associated with EVO bits—a wavelike motion occurring 3 to 4 s after the initial spike at the window’s center. This characteristic oscillation may serve as a unique indicator of the EVO fault. (ii) *DSTA1:WSF (Water Summary Fault)*: These faults appear in the right region of cluster 7 and are characterized by a sharp, rectangular drop of more than -100° lasting approximately 0.5 s. See part (b) of Fig. 6 for more details. (iii) *DSTA2:EVC (Modulator EOLC Fault—End of Line Clipper)*: These faults appear in parts of cluster 2 and cluster 5. In

the phase signal, they show up as a sharp rectangular jump starting in the middle of the time window and lasting about 1 s, with a magnitude of about 100° or more. After the jump, a wavelike motion follows for about 2 s. Refer to part (c) of Fig. 6 for more details. (iv) *DSTA1:RE (Klystron Reflected Energy Fault)*: As seen in parts of cluster 2 and cluster 5, these faults show up in similar regions to those of EVC Fault. Part (c) of Fig. 6 shows a sharp, rectangular jump starting midway through the time window, lasting about 1 s with a magnitude of around 100° or more. This is followed by a wavelike motion lasting about 2 s. It is common to see EVC and RE faults together, possibly caused by ringing in the klystron or pulse tank that feeds back into the PFN in the modulator.

To further investigate the status bits, we focused on three distinct regions, denoted as region A, region B, and region C, as shown in Fig. 5. A magnified view of these regions is provided in Fig. 11.

In region A, we observed that 85% of the anomalous samples had the DSTA2:EVO status bit activated. This finding reinforces a strong correlation between the physical signature in the phase fast signal within this region—specifically, a wavelike motion occurring 3 to 4 s after the initial spike at the center of the window—and the Modulator EVOC Fault, which corresponds to an Electrode Voltage Overcurrent.

In region B, we found that 8% and 2% of phase anomalies were associated with the DSTA2:EVC and DSTA1:RE status bits, respectively. While these percentages are relatively low, it is possible that the low detection rate reflects the low recall of these status bits. Although it is challenging to definitively conclude that the inactive status bits represent missed cases, we found that approximately 40% and 9% of the anomalies are associated with the DSTA2:EVC and DSTA1:RE status bits, respectively, under the following circumstances: the anomalies occurred in the same klystron and within a 24-h window of another anomaly where the status bit was active. Thus, anomalies within parts of cluster 2 and cluster 5 may be indicative of DSTA2:EVC or DSTA1:RE faults.

In region C, 16% of anomalies had the DSTA1:WSF status bit activated. Applying the same criteria as in the previous case, we found that approximately 60% of anomalies in this region occurred in the same klystron and within the same 24-h window as another anomaly where the WSF status bit was active. Notably, the majority of anomalies in this region were traced to station LI30:51. Additionally, data from the CATER system [30] at SLAC—a system where operators manually record faults and their resolutions—indicated that station 23:81 experienced a water leak fault, and the embeddings of its anomalous phase signals lie within region C.

As shown in Fig. 5, nearly all Modulator EVOC Faults are grouped in cluster 3, while Modulator EOLC Faults and Reflected Energy Faults are primarily associated with clusters 2 and 5. Similarly, Water Summary Faults are almost

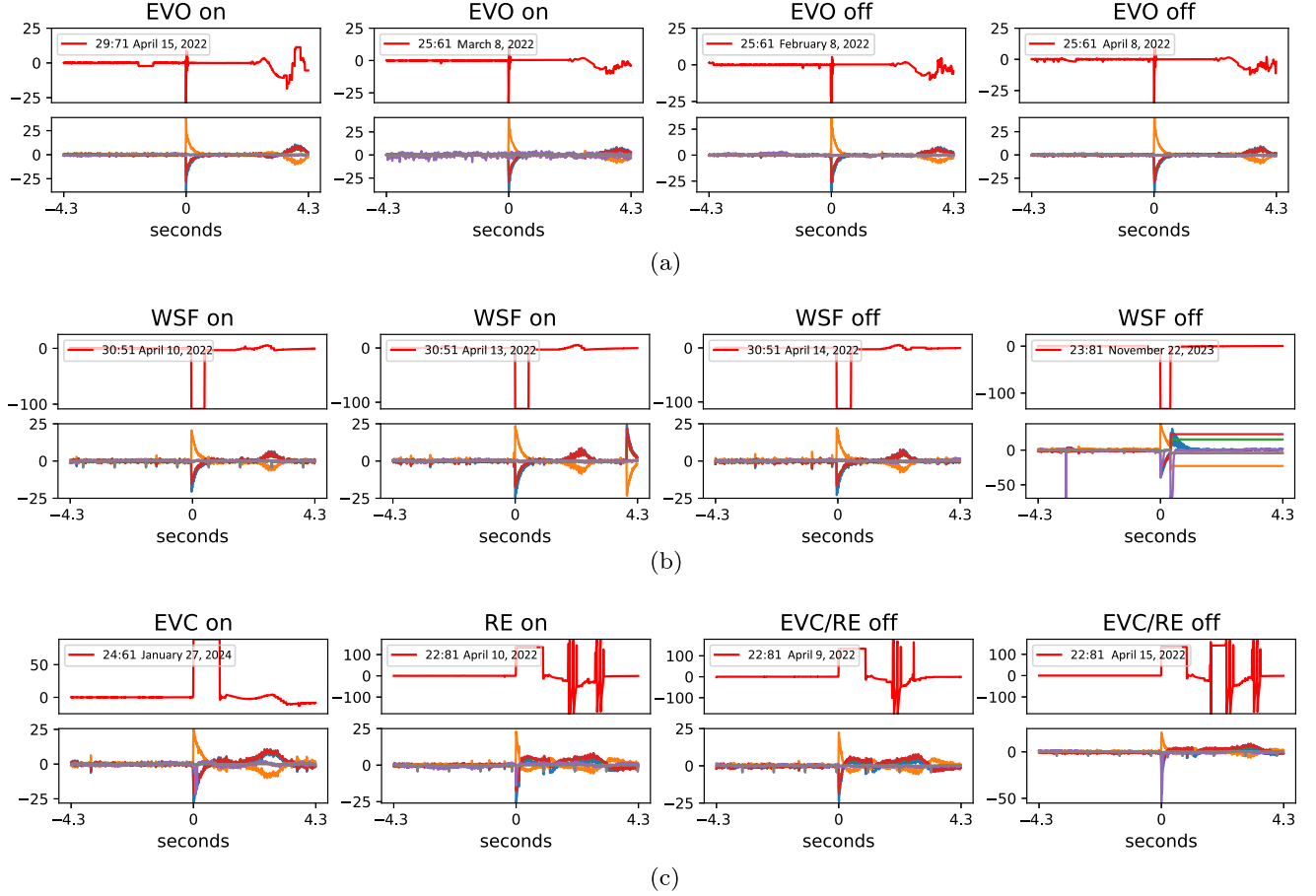


FIG. 6. Details of the phase signals associated with three categories of status bits from different clusters in Fig. 5: (a) EVO faults from cluster 3, (b) WSF faults from cluster 7 (b), and (c) combined EVC and RE faults from clusters 2 and 5. Four examples are shown for each category, with each example consisting of two components: the anomalous phase signals (shown in the first row) and the corresponding BPM signals (displayed in the second row). EVC and RE status bits are shown together due to similarity of behavior.

exclusively located in cluster 7. These associations suggest that the clustering results are indicative of underlying fault mechanisms. Moreover, each cluster exhibits distinct physical signatures in the phase data, further implying that phase patterns are reflective of the corresponding root causes. Notably, in the regions linked to the aforementioned clusters (regions A, B, and C), the relevant status bits are not always active. However, as illustrated in Fig. 6, samples with inactive status bits still display anomalous behaviors and phase signatures similar to those observed when the status bits are active. This observation suggests that status bits have limited recall, likely due to their lower sampling rate (0.2 Hz) relative to the phase data (120 Hz). Overall, phase data enable the detection of substantially more fault instances and provide physical signatures that are potentially diagnostic of root causes. Even when clusters contain faults associated with multiple status bits and distinct underlying mechanisms, they offer valuable operational insight by narrowing the scope of possible causes. The presence of active status bits in many cases further supports that phase-based anomalies correspond to genuine hardware issues.

D. Interpreting CoAD model with Shapley value

During the anomaly candidate identification step (Sec. IV A), the onset of the anomaly was positioned at the center of the time window. As a result, anomalies in the rf phase signal and BPM are expected to manifest at or near the center of the time window. Since most anomaly detection models provide only a single label for the entire time window, the precise timing of anomalies is not specified in the output.

Interpretable machine learning methods are designed to elucidate how models make predictions. In this context, an effective model should exhibit a strong focus on the center of each time window. This central focus should be evident in the results of interpretable methods. Therefore, to assess whether our CoAD algorithms are correctly attending to the relevant temporal regions during prediction, we employ Shapley values [31] to analyze model behavior. The Shapley value, a concept from cooperative game theory, allocates a fair share of the total contribution to each participant based on their individual inputs. For each pair of rf phase and BPM signals, we use the SHAP library [31] to

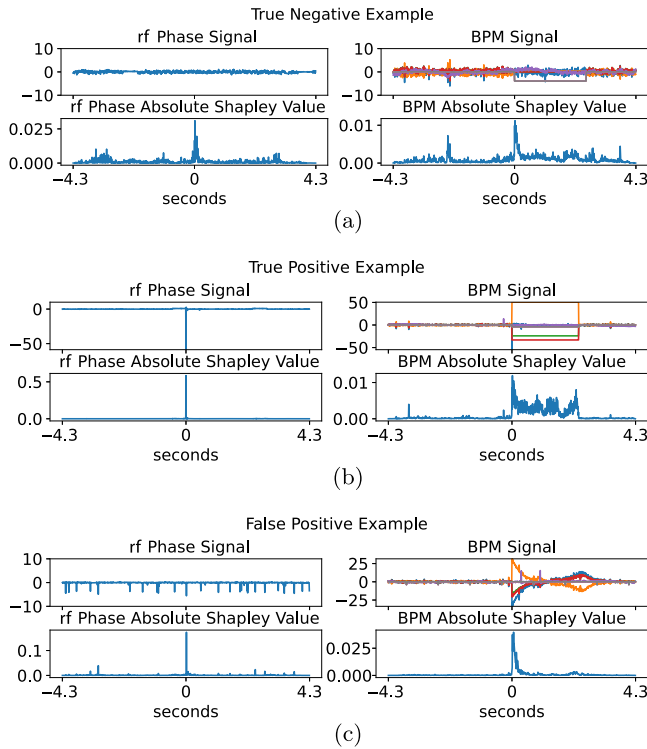


FIG. 7. Shapley value-based explanation of CoAD predictions. Subplots (a) and (b) show examples of correct predictions, with (c) showing a false positive. The model primarily focuses on the center of the time window, consistent with the candidate generation process, which attempts to place fault initiation at $t = 0$. In subplot (c), the model disregards broader statistical properties, such as the standard deviation of the time window, resulting in a false positive.

compute Shapley values during the inference process, assigning a single Shapley value to each timestamp. Subsequently, we compute the absolute values of the Shapley scores. Examples of a normal and an abnormal case are illustrated in parts (a) and (b) of Fig. 7, respectively. These figures demonstrate that the model predominantly attends to the center of the time window for anomaly predictions, which aligns with expected behavior. This observation indicates that the CoAD method performs effectively in focusing on the correct temporal regions for anomaly detection.

A more insightful application of Shapley values is to understand why the model makes incorrect predictions. As illustrated in part (c) of Fig. 7, there is an example where the model incorrectly classifies a normal sample as an anomaly. The sample is, in fact, normal because the rf phase spike at the center of the time window is part of the background noise in the rf phase, rather than indicative of anomalous behavior. This is evident from the presence of multiple spikes within the time window, suggesting that the rf station was experiencing consistent spikes during that period. Consequently, the spike at the center of the window is not due to a genuine anomaly in the rf phase. However,

Shapley value analysis reveals that the model primarily focuses on the center of the window when making predictions, while disregarding the spikes in other parts of the time window. This indicates that the model overlooks broader statistical properties, such as the standard deviation of the time window, and instead singularly concentrates on the central region. This insight highlights a potential improvement for CoAD, suggesting that incorporating the standard deviation of the time window in the rf phase could enhance the model's performance in distinguishing true anomalies from background noise.

VI. CONCLUSION

The inherent complexity of rf phase data has historically hindered its application to anomaly detection at SLAC. While classical methods fail to extract value from phase data, in this work we use the CoAD framework to train neural networks that operate on phase data. We show that deep learning models operating on phase data detect almost 3 times as many anomalies compared to models using amplitude data alone. We also find that CoAD outperforms existing traditional and deep learning methods. Moreover, the additional anomalies detected through phase data significantly increase the coverage of rf stations compared to amplitude-based detection, particularly in the critical injector region.

The rich information in phase data suggests that in addition to indicating the occurrence of anomalies, they can also provide more granularity in the prediction of the root cause of failure. We apply dimensionality reduction and clustering methods to identify distinct groups within the anomalous events identified by CoAD. By integrating data from SLAC's CATER system and status bits, we show that some individual clusters could correlate with specific rf station failure mechanisms, providing additional information to operators to support recovery from failures.

We also investigated the interpretability of the networks using Shapley values. Through these values, we observed that each neural network within the CoAD framework focused on the center of its respective time window, aligning with expectations based on the data generation process. These findings affirm that our CoAD model is functioning as intended and also provide insight into situations in which CoAD makes mistakes, informing future avenues of research. In future work, we plan to investigate the detection of repeated spike faults in the phase data, which currently remain undetectable by CoAD. Additionally, while CoAD presently operates on two data streams, we aim to extend its framework to accommodate multiple streams by leveraging graph neural networks. More broadly, the successful application of CoAD to this complex system highlights its potential for scalable anomaly detection across the thousands of subsystems within linear accelerators.

We are in the process of deploying the CoAD models within the operational environment at SLAC. For this deployment, a model trained offline is integrated into an online inference framework. The end-to-end pipeline consists of three primary processes: the first continuously collects real-time system data; the second cleans and pre-processes the acquired data while identifying potential anomaly candidates; and the third applies the trained CoAD model to determine these candidates as either true anomalies or normal behavior. When a candidate is confirmed as a true anomaly in the third stage, an alert is automatically issued to the operator via a graphical user interface. As discussed in recent work [32], continual learning is critical for sustaining long-term deployment performance in online settings. We are expanding the training dataset to include several years of data to capture the widest range of anomalies. To mitigate potential data drift, we plan to periodically retrain the model offline using newly collected data and then redeploy the updated model to the online system.

ACKNOWLEDGMENTS

We thank Jim Craft, Xupeng Chen, Howard V. Smith, Benjamin Ripman, and Timothy Maxwell for their helpful discussions. The use of the LINAC Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515.

DATA AVAILABILITY

Some of the data that support the findings of this article are openly available [24]. Dataset used in Secs. VA and VC are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX A: CHALLENGING DETECTION SCENARIOS SUCCESSFULLY HANDLED BY COAD

In Fig. 8, we present two examples of subtle anomalies successfully detected by CoAD. These cases are challenging because the rf phase exhibits only a minor jump at the center of the window, confined to a single pulse. The anomaly is confirmed by a corresponding jump in the BPM signals at the same position.

In Fig. 9, we present two examples of challenging normal cases that CoAD classifies correctly. The first case [Fig. 9(a)] exhibits wavelike motion in the rf phase at the center of the window, caused by the system's periodic phase-scan operation performed every few minutes. Any coincident fluctuations in the BPM signals during this

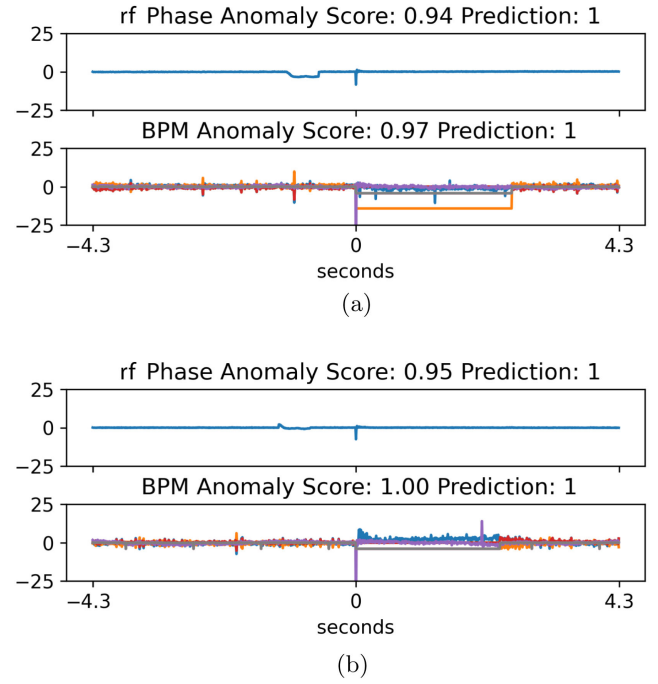


FIG. 8. Two examples (a) and (b) each showing CoAD detecting challenging cases in which the rf phase exhibits only a minor jump at the center of the time window.

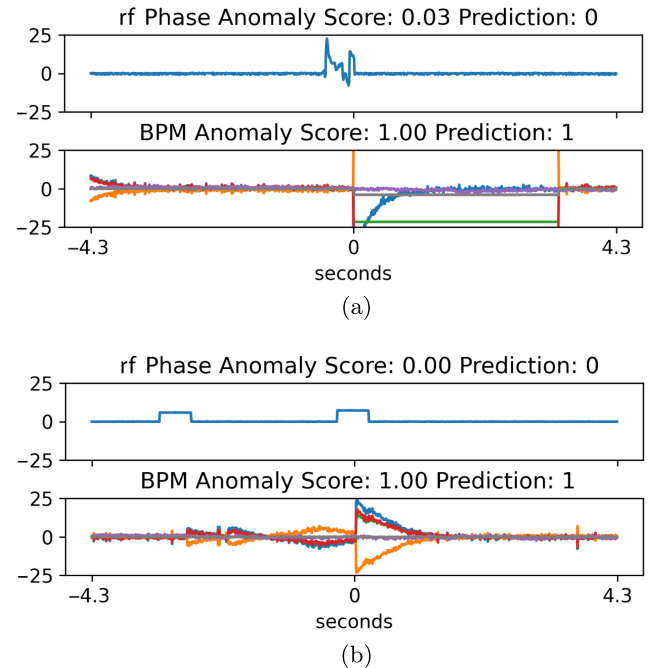


FIG. 9. (a) CoAD correctly identifies the wavelike motion in the rf phase near the center of the window as normal behavior, since it arises from the system's periodic phase-scan operation performed every few minutes. (b) Similarly, CoAD correctly classifies the periodic rectangular waveforms in the rf phase as normal because they are not synchronous with the BPM faults.

interval arise from independent processes unrelated to the rf stations and therefore do not indicate an rf-induced fault. Having encountered numerous phase-scan patterns during training, CoAD reliably labels such instances as normal. Notably, in most cases, these phase-scan patterns correspond exclusively to normal BPM behavior, enabling the neural network trained with the coincidence-based loss to recognize this waveform as normal. The second case [Fig. 9(b)] contains rectangular waveforms in the rf phase, which also occur periodically and are not responsible for BPM faults. CoAD likewise identifies these as normal, recognizing that such patterns do not induce anomalous BPM behavior.

Traditional anomaly detection methods, such as fixed-threshold and statistical outlier, often struggle with both the subtle anomalies in Fig. 8 and the challenging normal cases in Fig. 9. In the anomaly cases, the rf phase change is small and confined to a single pulse, falling below typical thresholds and being smoothed out by longer temporal windows, leading to missed detections. Reducing thresholds to capture such events, however, would cause a surge in false positives from benign fluctuations. Conversely, in the normal cases, benign but unusual patterns—such as wavelike motion from phase-scan operations or periodic rectangular waveforms—deviate from the nominal baseline and may coincide with BPM variations caused by unrelated processes. Traditional methods may misinterpret these correlations as evidence of faults. These challenges are exacerbated by the sensitivity of such methods to hyperparameters, including thresholds, window sizes, and the number of consecutive anomalous points required for detection. A single global setting cannot simultaneously capture weak, short-duration anomalies while suppressing benign but rare operational patterns. In contrast, CoAD’s coincidence-based learning approach uses training data to identify cross-signal patterns that truly indicate faults, enabling it to distinguish subtle anomalies from normal cases without reliance on brittle, globally tuned thresholds.

APPENDIX B: REPEATED SPIKE FAULT

Using synchronous rf phase diagnostic data, we identified a type of event that would be overlooked when only considering rf amplitude data. We refer to this special event as the repeated spike fault. These events are visualized in Fig. 10. As shown, these faults are sufficiently small that

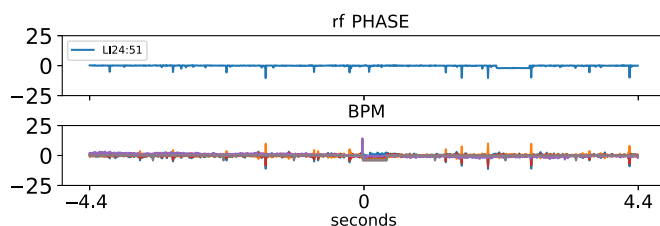


FIG. 10. An example of repeated spikes fault.

they do not exceed the BPM candidate-selection threshold, meaning that they are not considered candidates for fault detection in the first step of our method, shown in Fig. 2. This is also why those repeated spikes do not appear at the center of each time window.

However, despite their small magnitude, repeated spike faults can occur at high frequency, still degrade beam quality, and are therefore undesirable. Consultation with domain experts revealed that these faults may result from thyatron ranging issues. In practice, such spike faults are generally considered tolerable provided that each spike occurs at intervals of no less than 23 s. As illustrated, these faults can repeat multiple times within each 8-s window, occurring simultaneously in both the BPM signals and the rf phase signals. Our investigation revealed that these faults primarily originated from rf stations LI:24:51 and LI:22:71 during the period from January 18, 2024 to January 30, 2024. Specifically, most of the faults from LI:24:51 are concentrated on January 23, 2024, while the majority of faults from LI:22:71 are concentrated on January 29, 2024. We estimate that the repeated spike fault occurred more than 10,000 times in LI:24:51 on January 23, 2024, and more than 2000 times in LI:22:71 on January 29, 2024.

The current candidate generation procedure described in Sec. IV A excludes repeated spike faults, as these faults typically do not surpass the threshold required for candidate selection. Consequently, these repeated spike faults often do not appear at the center of the data window and are labeled as normal samples during the labeling process. Since our primary focus is on identifying hard faults rather than persistent instability, we refrain from adjusting the threshold to capture these events. Reducing the threshold to account for persistent instability would significantly increase the number of candidate samples, as repeated spike faults can occur thousands of times within a single day.

APPENDIX C: ENLARGED VIEW OF FIG. 5

Figure 11 presents an enlarged view of regions A, B, and C from Fig. 5, each corresponding to distinct fault types: region A captures Modulator EVOC Faults (EVO), region B includes both Modulator EOLC Faults (EVC) and Reflected Energy Faults (RE), and region C corresponds to Water Summary Faults (WSF). Within each region, we highlight several representative examples with status bits both activated (on) and deactivated (off).

A key observation across all regions is that anomalous phase signatures appear similar regardless of the status bit state. For example, in region A, both on- and off-bit cases exhibit a distinctive wavelike motion in the phase fast signal, occurring approximately 3–4 s after a sharp spike centered in the window. Additionally, these on-bit and off-bit cases are located near each other in the UMAP embeddings. This spatial proximity, combined with the similarity of their phase signatures, suggests that the off-bit cases likely represent true anomalies that were missed by

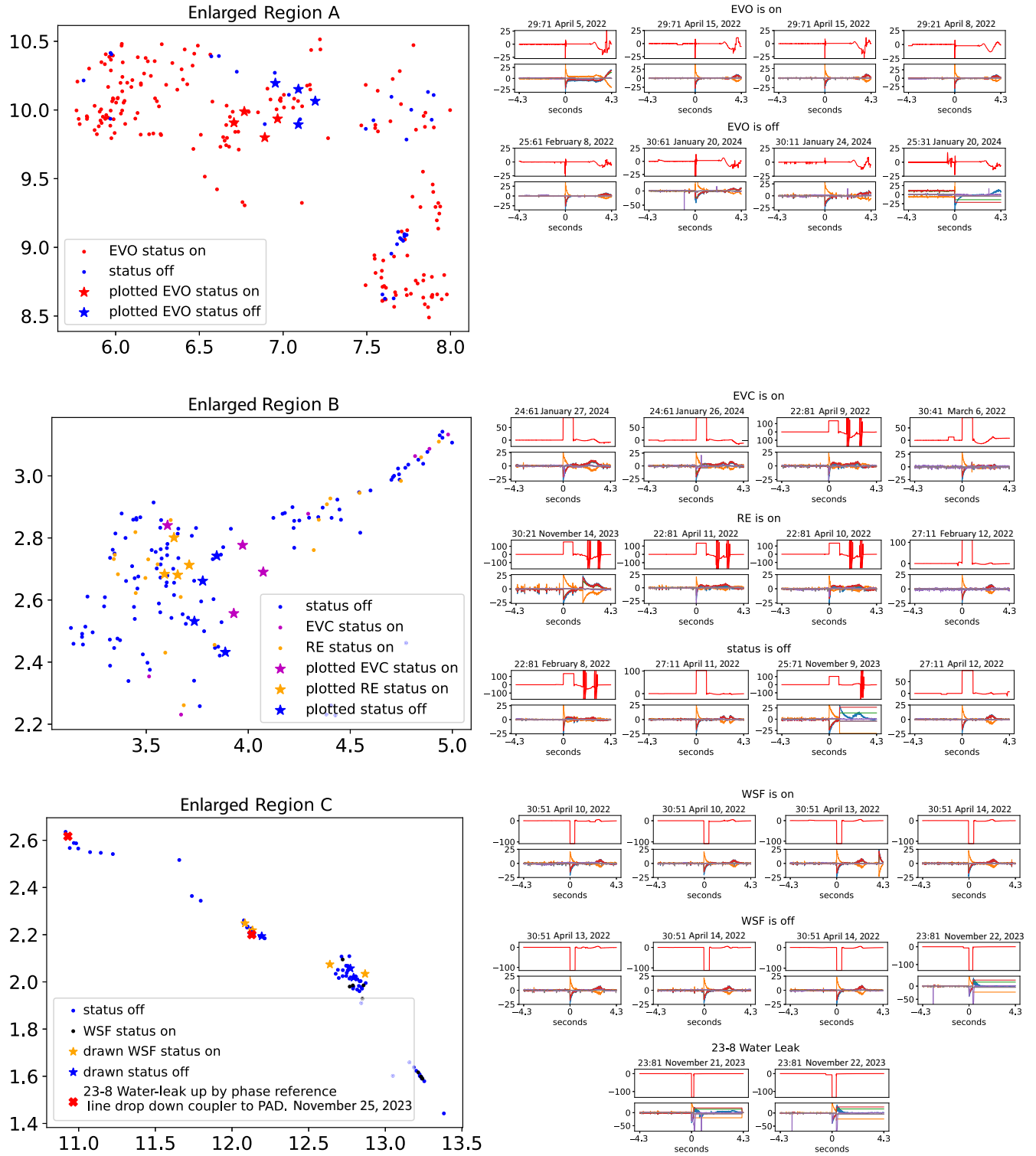


FIG. 11. Enlarged view of Fig. 5 (at left) along with phase and BPM data for select examples (at right). Each example consists of two components: the anomalous phase signals (shown in the first row) and the corresponding BPM signals (displayed in the second row). In enlarged region A, 85% of the examples have EVO status bits as ON. In enlarged region B, 8% and 2% of RE status bits and EVC status bits, respectively, are ON, highlighting the poor recall of some status bits. EVC and RE examples are plotted together given the similarity of phase data in examples corresponding to those status bits. In enlarged region C, 16% of the WSF status bits are ON.

the status bits but identified through the phase-based analysis. This, in turn, implies that the status bits have low recall. Overall, these findings indicate that phase-based anomaly detection identifies genuine faults that are sometimes missed by status bits. The consistency of the physical signatures—regardless of status bit activation—and their alignment with known fault types indicate that these anomalies correspond to real underlying issues in the rf system.

APPENDIX D: CANDIDATE SELECTION

This section outlines the procedure for selecting the most anomalous rf stations as part of the candidate generation step described in Sec. IV A. Note that this method is specifically tailored to our dataset and includes certain preprocessing steps.

To identify the most anomalous klystron stations, we begin by computing the absolute phase values for each rf station at every timestamp. Let $\Phi \in \mathbb{R}^{N \times T}$ denote the phase matrix, where N is the number of rf stations and T is the total number of time steps. At each time $t \in \{1, \dots, T\}$, the absolute phase for station $n \in \{1, \dots, N\}$ is computed as

$$\phi_{n,t} = |\Phi_{n,t}|,$$

where $\Phi_{n,t}$ represents the original (signed) phase value. The absolute value is used because we are interested in the magnitude of deviation from zero, regardless of sign.

To reduce the inclusion of system-level faults—characterized by uniform increases in phase values across multiple stations—we apply a preprocessing step that subtracts the fifth-largest absolute phase value at each time point. Specifically, let $\phi_i^{(5)}$ denote the fifth-largest value among $\{\phi_{1,t}, \dots, \phi_{N,t}\}$. The adjusted phase for station n at time t is then given by

$$\tilde{\phi}_{n,t} = \phi_{n,t} - \phi_i^{(5)}.$$

Next, for each station n , we define a station-specific time window $(t_i^n - g, t_i^n]$, where t_i^n is the trigger time associated with station n , and g is the number of preceding pulses to consider. The value of g is set to 20, the same as the value of k , the number of consecutive pulses used to aggregate to calculate the final anomaly score. This window is intended to capture the earliest onset of anomalous behavior. We are confident that the earliest onset of anomalous behavior will occur within $(t_i^n - k, t_i^n)$. For this reason, we set g equal to k . Within this window, we compute the maximum adjusted phase:

$$d_n = \max_{t \in (t_i^n - g, t_i^n]} \tilde{\phi}_{n,t}.$$

We then identify the five stations with the largest values of d_n and designate them as anomaly candidates. From this set, we retain only those stations for which $d_n > \tau$, where

Algorithm 1. Most anomalous klystron station identification (station-specific trigger times).

Output: Set of selected anomalous RF stations $\mathcal{A}$

```

1 for each time  $t = 1$  to  $T$  do
2   Compute absolute phases:  $\phi_{n,t} \leftarrow |\Phi_{n,t}|$  for all
    $n \in [1, N]$ ;
3   Find the 5th-largest phase across all  $n$ :  $\phi_t^{(5)}$ ;
4   for each station  $n = 1$  to  $N$  do
5     Adjusted phase:  $\tilde{\phi}_{n,t} \leftarrow \phi_{n,t} - \phi_t^{(5)}$ ;
6 for each station  $n = 1$  to  $N$  do
7   Define station-specific window:
    $\mathcal{W}_n \leftarrow \{\tilde{\phi}_{n,t} \mid t \in (t_i^n - g, t_i^n]\}$ ;
8   Compute max deviation:  $d_n \leftarrow \max(\mathcal{W}_n)$ ;
9  $\mathcal{C} \leftarrow$  indices of top 5 stations with largest  $d_n$ ;
10  $\mathcal{A} \leftarrow \{n \in \mathcal{C} \mid d_n > \tau\}$ ;
11 if  $\mathcal{A} = \emptyset$  then
12    $n^* \leftarrow \arg \max_{n \in \mathcal{C}} d_n$ ;
13    $\mathcal{A} \leftarrow \{n^*\}$ ;
14 return  $\mathcal{A}$ ;
```

$\tau = 25$ is a predefined threshold. τ is set to 25 because the majority of the genuine anomalous phase behavior will exceed this threshold. If none of the five candidates exceeds this threshold, we instead select the single station with the largest d_n value.

If only one station is selected, its rf signal is paired with the corresponding beam data to form a single candidate instance. If multiple stations exceed the threshold, each selected rf signal is paired with the same beam data, resulting in multiple candidate instances for further analysis. Formally, let us denote the rf diagnostic data as s , describing the subsystem, and the BPM data as q , representing the overall quality of the system. Consequently, each candidate C_i is defined as $C_i = \{s_i, q_i\}$. The method of identifying the most anomalous klystron station(s) is summarized in Algorithm 1.

APPENDIX E: COAD TRAINING DETAILS

We use the dataset constructed in Sec. IV A with additional transformations: (i) normalizing the data, (ii) adding $\mathcal{N}(0, 0.1)$ noise to each input, and (iii) randomly rolling the BPM data in time between 0 and 10 points. We define our models A_{θ_s} and A_{θ_l} as CNNs. As shown in Fig. 12(a), our s network consists of two 1D convolutions with one output channel, kernel size 20, padding 0, and stride 5, followed by three fully connected layers with output sizes 6, 6, and 1. As shown in Fig. 12(b), our q network has the same structure. All layers except the final layer are followed by a ReLU activation. We train for 2000 epochs with a batch size of 400, using the Adam optimizer with a learning rate of 10^{-4} . During each epoch, we randomly hold out 15% of the training data for validation and perform early stopping based on the $\hat{F}_\beta$ metric on the validation set.

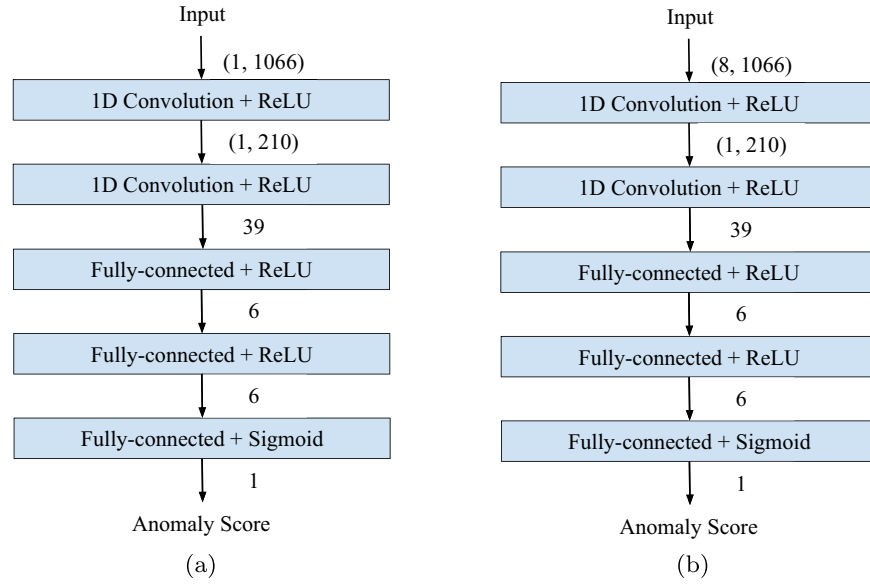


FIG. 12. Model architecture for the particle accelerator DNNs, showing the input size to each layer. (a) Architecture for rf station phase data and (b) architecture for BPM data.

APPENDIX F: CANDIDATE GENERATION HYPERPARAMETERS

Please refer to Table V for the list of hyperparameters used in the candidate generation process, along with their values and the rationale for each choice.

APPENDIX G: DETAILED EXPLANATION OF THE STATUS BITS

The following are detailed explanations of the status bit mentioned in Sec. VC: (i) *DSTA2:EVO (Modulator EVOC Fault—Electrode Voltage Overcurrent)*: The klystron is

powered by a high-voltage pulse supplied by the modulator. This pulse is increased to the needed level by a transformer. The voltage and current of this pulse are monitored by sensors and reported to the control system. If the current in the klystron pulse is too high, a safety system called the EVO interlock will shut down the modulator to prevent damage. In older modulators, a series of high-current signals can indicate a problem inside the klystron's pulse tank. This could mean that the modulator is producing a higher voltage than it should. By checking these signals on an oscilloscope (a device that shows electrical signals overtime), you can see exactly when and where the issue

TABLE V. Hyperparameter settings for candidate generation.

Hyperparameter	Value	Purpose
l , lagging rolling window size	600	Corresponds to ~ 5 s. Chosen to be longer than short-duration dips (where the machine recovers quickly) to prevent them from disproportionately affecting the median, yet shorter than the typical time scale of machine drift.
n , number of signals	8	Four BPMs, each providing either a TMIT or an X or Y position measurement, for a total of eight distinct signals. These BPMs are placed in regions with nonzero horizontal or vertical dispersion, making them sensitive to beam energy variations, and are located downstream of all rf stations to capture cumulative upstream effects.
k , consecutive pulses for final score	20	Favors detection of longer-duration anomalies. Values such as 30 or 40 also capture sustained events, but excessively large values may exclude many true anomalies.
t , threshold for candidate determination	8	Higher values yield a higher concentration of anomalies in the candidate set. A value of 8 balances concentration with coverage; very large values may exclude many true anomalies. Similar values (10, 12, 15) also perform effectively.
w , candidate window size	533	Corresponds to ~ 4.3 s. Phase anomalies and fault recovery typically last several seconds. Increasing the window beyond this length provides minimal benefit.

occurs. A sudden spike in the signal could be caused by an electrical arc (a high-voltage discharge) somewhere in the high-voltage system, which can also trigger a fault. (ii) *DSTA1:WSF (Water Summary Fault)*: happens when any of the various flow switches that monitor cooling water to the station are faulted. (iii) *DSTA2:EVC (Modulator EOLC Fault—End of Line Clipper)*: happens when too much current flows due to high reverse voltage in the system. A small amount of reverse voltage helps the thyatron switch off properly, but if it exceeds 5–6 kV, it can damage components. A protective circuit (thyrite stack) limits this voltage but increases current, which can trigger the EOLC fault and shut down the modulator to prevent damage. (iv) *DSTA1:RE (Klystron Reflected Energy Fault)*: happens when too much energy is reflected back into the klystron instead of being transferred to the load (like a particle accelerator or radar system). Klystrons are designed to amplify and transmit high-power microwave signals, but if the signal is not absorbed properly by the load, the energy can reflect back into the klystron and cause damage or instability. Common causes include load mismatches, waveguide arcing, vacuum leaks, power overload, and component misalignment or failure. Protective systems detect these faults and shut down the modulator to prevent damage.

-
- [1] Ryan Humble, Finn H. O'Shea, William Colocho, Matt Gibbs, Helen Chaffee, Eric Darve, and Daniel Ratner, Beam-based rf station fault identification at the SLAC Linac Coherent Light Source, *Phys. Rev. Accel. Beams* **25**, 122804 (2022).
 - [2] Ryan Humble, Zhe Zhang, Finn O'Shea, Eric Darve, and Daniel Ratner, Coincident learning for unsupervised anomaly detection of scientific instruments, *Mach. Learn.* **5**, 035036 (2024).
 - [3] Zhiyuan Chen, Wei Lu, Radhika Bhong, Yimin Hu, Brian Freeman, and Adam Carpenter, Anomaly detection of particle orbit in accelerator using LSTM deep learning technology, *arXiv:2401.15543*.
 - [4] Chris Tennant, Brian Freeman, Reza Kazimi, Daniel Moser, Dan Abell, Jonathan Edelen, and Joshua Einstein-Curtis, A smart alarm for particle accelerator beamline operations, *Mach. Learn.* **4**, 015021 (2023).
 - [5] Davide Marcato, Damiano Bortolato, Valentina Martinelli, Giovanni Savarese, and Gian Antonio Susto, Time-series deep learning anomaly detection for particle accelerators, *IFAC-PapersOnLine* **56**, 1566 (2023).
 - [6] Willem Blokland, Kishansingh Rajput, Malachi Schram, Torri Jeske, Pradeep Ramuhalli, Charles Peters, Yigit Yucesan, and Alexander Zhukov, Uncertainty aware anomaly detection to predict errant beam pulses in the Oak Ridge Spallation Neutron Source accelerator, *Phys. Rev. Accel. Beams* **25**, 122802 (2022).
 - [7] Ihar Lobach and Michael Borland, Long short-term memory networks for anomaly detection in magnet power supplies of particle accelerators, *arXiv:2405.18321*.
 - [8] Yusuke Suetsugu, Machine-learning-based pressure-anomaly detection system for SuperKEKB accelerator, *Phys. Rev. Accel. Beams* **27**, 063201 (2024).
 - [9] Kishansingh Rajput, Malachi Schram, Willem Blokland, Yasir Alanazi, Pradeep Ramuhalli, Alexander Zhukov, Charles Peters, and Ricardo Vilalta, Robust errant beam prognostics with conditional modeling for particle accelerators, *Mach. Learn.* **5**, 015044 (2024).
 - [10] Majdi I. Radaideh, Chris Pappas, Mark Wezensky, Pradeep Ramuhalli, and Sarah Cousineau, Early fault detection in particle accelerator power electronics using ensemble learning, *Int. J. Prognostics Health Manage.* **14** (2023).
 - [11] Md M. Rahman, Adam Carpenter, Khan Iftekharruddin, and Chris Tennant, Accelerating cavity fault prediction using deep learning at Jefferson Laboratory, *Mach. Learn.* **5**, 035078 (2024).
 - [12] Zhou Jiqing, Li Deming, and Su Haijun, Particle accelerator power system early fault diagnosis based on deep learning and multi-sensor feature fusion, *Eng. Res. Express* **6**, 025225 (2024).
 - [13] Zhe Yang, Rongbing Ye, Lingli Jiang, Jianyu Long, Yunwei Huang, and Chuan Li, Few-shot fault diagnosis of particle accelerator power system using a bidirectional discriminative prototype network, *Appl. Soft Comput.* **170**, 112726 (2025).
 - [14] Yigit A. Yucesan, Willem Blokland, Pradeep Ramuhalli, Alexander Zhukov, Charles Peters, David Brown, and Cary Long, A machine learning approach for particle accelerator errant beam prediction using spatial phase deviation, *Nucl. Instrum. Methods Phys. Res., Sect. A* **1063**, 169232 (2024).
 - [15] Jonathan P. Edelen and Nathan M. Cook, Anomaly detection in particle accelerators using autoencoders, *arXiv:2112.07793*.
 - [16] T. Straumann, R. Akre, R. G. Johnson, D. Kotturi, P. Krejcik, E. Medvedko, J. Olsen, and S. Smith, LCLS beam-position monitor data acquisition system, in *Proceedings of the ICALEPCS07, Knoxville, TN* (Joint Accelerator Conference Website, CERN, Geneva, Switzerland, 2007).
 - [17] Ryan Prescott Adams and David J. C. MacKay, Bayesian online changepoint detection, *arXiv:0710.3742*.
 - [18] Natalia Y. Bilenko and Jack L. Gallant, Pyreca: Regularized kernel canonical correlation analysis in Python and its applications to neuroimaging, *Front. Neuroinf.* **10**, 49 (2016).
 - [19] Eun Jeong Min, Eric C. Chi, and Hua Zhou, Tensor canonical correlation analysis, *Stat* **8**, e253 (2019).
 - [20] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou, Isolation forest, in *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining* (IEEE, New York, NY, 2008), pp. 413–422.
 - [21] Larry M. Manevitz and Malik Yousef, One-class SVMs for document classification, *J. Mach. Learn. Res.* **2**, 139 (2001), <https://www.jmlr.org/papers/v2/manevitz01a.html>.
 - [22] Cristian I. Challu, Peihong Jiang, Ying Nian Wu, and Laurent Callot, Deep generative model with hierarchical latent factors for time series anomaly detection, in *Proceedings of the International Conference on Artificial Intelligence and Statistics* (PMLR, Cambridge, MA, 2022), pp. 1643–1654.

- [23] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei, Robust anomaly detection for multivariate time series through stochastic recurrent neural network, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Association for Computing Machinery, New York, NY, 2019), pp. 2828–2837.
- [24] J. Liang *et al.*, LCLS RF Station Phase Anomaly Candidate Dataset. SLAC National Accelerator Laboratory (SLAC), Menlo Park, CA (United States), 2025, [10.71929/2584368](https://doi.org/10.71929/2584368).
- [25] Navid Hashemi, Eduardo Verdugo German, Jonatan Pena Ramirez, and Justin Ruths, Filtering approaches for dealing with noise in anomaly detection, in *Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC)* (IEEE, New York, NY, 2019), pp. 5356–5361.
- [26] Yue Zhao, Ryan A. Rossi, and Leman Akoglu, Automating outlier detection via meta-learning, [arXiv:2009.10606](https://arxiv.org/abs/2009.10606).
- [27] Willem van Veluw, Application of mixture models to threshold anomaly scores, Master's thesis, Utrecht University, 2023.
- [28] Leland McInnes, John Healy, and James Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, [arXiv:1802.03426](https://arxiv.org/abs/1802.03426).
- [29] Leland McInnes, John Healy, and Steve Astels, HDBSCAN: Hierarchical density based clustering, *J. Open Source Software* **2**, 205 (2017).
- [30] D. Rogind, W. Allen, W. Colacho, G. DeContreras, J. Gordon, P. Pandey, and H. Shoaee, Integrated management tool for controls software problems, requests and project tasking at SLAC, in *Proceedings of the ICALEPCS* (Joint Accelerator Conference Website, CERN, Geneva, Switzerland, 2018).
- [31] Scott M. Lundberg and Su-In Lee, A unified approach to interpreting model predictions, in *Advances in Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2017), Vol. 30.
- [32] Kishansingh Rajput, Sen Lin, Auralee Edelen, Willem Blokland, and Malachi Schram, Outlook towards deployable continual learning for particle accelerators, *Mach. Learn.* **6**, 031001 (2025).