Data in Brief 51 (2023) 109658

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Operational data for fault prognosis in particle accelerators with machine learning

Majdi I. Radaideh[a,b,*], Chris Pappas[a], Mark Wezensky[a], Sarah Cousineau[a]

[a] Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, TN 37830, United States
[b] Department of Nuclear Engineering and Radiological Science, University of Michigan, Ann Arbor, MI 48109, United States

ARTICLE INFO

Article history:
Received 12 June 2023
Revised 29 September 2023
Accepted 3 October 2023
Available online 11 October 2023

Dataset link: Operational Data for Fault Prognosis in Particle Accelerators with Machine Learning (Original data)

Keywords:
Fault prognosis
Health management
Machine learning
Signal processing
Spallation neutron source
High voltage converter modulators

ABSTRACT

This paper presents real operational data collected from the power systems of the Spallation Neutron Source facility, which provides the most intense neutron beam in the world. The authors have used a radio-frequency test facility (RFTF) and simulated system failures in the lab without causing a catastrophic system failure. Waveform signals have been collected from the RFTF normal operation as well as during fault induction efforts. The dataset provides a significant amount of normal and faulty signals for the training of statistical or machine learning models. Then, the authors performed 21 test experiments, where the faults are slowly induced into the RFTF system for the purpose of testing the models in fault prognosis to detect and prevent impending faults. The test experiments include interesting combinations of magnetic flux compensation and start pulse width adjustments, which cause gradual deterioration in the waveforms (e.g., system output voltage, system output current, insulated-gate bipolar transistor currents, magnetic fluxes), which mimic the fault scenarios. Accordingly, this dataset can be valuable for developing models to predict impending fault scenarios in power systems in general and in particle accelerators in specific. All experiments occurred in the Spallation Neutron Source facility of Oak Ridge National Laboratory in Oak Ridge, Tennessee of the United States in July 2022.

* Corresponding author.
E-mail address: radaideh@umich.edu (M.I. Radaideh).

https://doi.org/10.1016/j.dib.2023.109658

Specifications Table

Subject	Electrical and Electronic Engineering
Specific subject area	Fault prognosis, machine learning, control engineering
Type of data	Table (time series)
How the data were acquired	The radio-frequency test facility (RFTF) data acquisition system records 12 waveforms with time length of 1.5 ms and a sampling rate of 400 ns, then writes them to an external hard drive. The data are structured in 3D arrays and saved to binary NumPy files.
Data format	Raw
Description of data collection	The raw data from the controller are reported without preprocessing. Each pulse in the dataset has a time length of 1.5 ms. The dataset includes a prognosis test set which involves a set of 21 experiments where fault precursors are induced into the system. Depending on the machine condition, the recorded pulse could be a normal or a faulty pulse.
Data source location	Facility: Spallation Neutron Source Institution: Oak Ridge National Laboratory Address: 8600 Spallation Dr, Oak Ridge, TN 37,830, United States
Data accessibility	The data are presented in a Mendeley data repository with the following information: • Title of the repository: Operational Data for Fault Prognosis in Particle Accelerators with Machine Learning • Mendeley Data: https://data.mendeley.com/datasets/9zxrt6pf2k • DOI:https://doi.org/10.17632/9zxrt6pf2k
Related Research Article	M. I. Radaideh, C. Pappas, M. Wezensky, P. Ramuhalli, S. Cousineau (2023). Early Fault Detection in Particle Accelerator Power Electronics Using Ensemble Learning. International Journal of Prognostics and Health Management, 14 (1), pp. 1–19.

1. Value of the Data

- This dataset provides high-quality signal data from the operation of the radio-frequency test facility in the spallation neutron source. This dataset is valuable to develop and test novel algorithms for fault prognosis, health management, and autonomous control in complex engineering systems like particle accelerators.
- Beneficiaries of this dataset include researchers, control, electronics, and maintenance engineers interested in machine learning, signal processing, predictive maintenance, and particle accelerators.
- This dataset can be used to design methods and algorthims to detect system faults ahead of time allowing operator intervention and proper maintenance management.

2. Objective

The Spallation Neutron Source (SNS) is a complex engineering system used to accelerate charged particles. The power systems at the SNS continuously suffer from operational anomalies and catastrophic failures that increase facility downtime. This dataset presents real operational data collected from the power systems of the SNS, including waveform signals collected from a radio-frequency test facility during normal and fault operational scenarios. The dataset provides valuable information for training statistical or machine learning models to predict impending

fault scenarios in power systems. This dataset is related to the research paper published in the International Journal of Prognostics and Health Management (see **Related Research Article** in the specifications table). The article highlights using machine learning models for early fault detection in particle accelerators. The study explores the performance of various machine learning models in the same experimental setup. This data article unveils the data that were used to train and test all models. Using this data article, the readers will be able to reproduce our results in the related research article and propose new statistical and machine learning methods that can help in early fault detection or prognosis applications in complex engineering systems like particle accelerators.

3. Data Description

High Voltage Converter Modulators (HVCM) provide power to the particle accelerator klystrons to produce high-power radio frequency to accelerate the particles in the spallation neutron source (SNS) [1]. HVCM systems are used to convert 3-phase 13.8 kVAC into a train of up to 74–135 kV, 1.3 ms pulses at 60 Hz to 92 klystrons. An HVCM consists of (1) AC switch gear and magnetics, (2) a six-pulse phase-controlled rectifier unit, (3) insulated-gate bipolar transistor (IGBT) devices configured for a polyphase, isolated, DC-DC converter to achieve a high voltage output, and (4) a controller rack [2].

Given the reliability issues and failures of the HVCM systems [3], the SNS continues to experience a significant downtime due to these HVCM failures. This motivates the authors to collect real-time operational data and to develop diagnosis and prognosis methods to help improving HVCM system reliability by detecting failures ahead of time and help preventing them. HVCMs are well instrumented to collect waveform data, and in this dataset, we established a test facility called Radio-Frequency Test Facility (RFTF) that runs in similar conditions as the linear accelerator SNS modulators, making the current experimental setup similar to operational HVCM systems at the SNS. The nature of this data is waveform/signal data collected from the RFTF and can be valuable for fault prognosis applications since they involve real-world and real-time fault detection scenarios. More details about the facility and the experimental setup can be found in the next section.

The dataset repository has a total of 48 files, classified into (1) *primary files* and (2) *supplementary files*. The primary files (24 files) include the NumPy binary data files and the CSV file that has the test labels, which are all what the user needs to proceed in the analysis. The supplementary files (24 files) are only to aid in interpretation and not required for the analysis. The supplementary files include excel files, png plots, and a Python script. *It is important to note that this classification serves as a reference and does not influence the organization of data within the repository*. In other words, both primary and supplementary files can coexist within the same directory. The purpose of this classification is to assist users in distinguishing between files essential for analysis and those provided solely to aid in interpretation. The data is structured as follows:

Main directory: **“data”**

- Subfile: “**load_dataset.py**”, see Table 1 for the description of this file.
- Subdirectory: “**train**”, see Table 2 for the description of the files in this subdirectory.
- Subdirectory: “**test**”, see Table 3 for the description of the files in this subdirectory.

Table 1
Description of the dataset files under the main directory “data”.

File	Type	Description	*Shape*
load_dataset.py*	Python script	A simple python script to load and plot the dataset	NA

* Supplementary file.

Table 2
Description of the dataset files under the subdirectory "train".

File	Type	Description	*Shape*
normal.npy	3D NumPy array (binary)	Array of 5000 **normal** pulses, each with 3753 time steps, and has 12 unique waveforms.	(5000, 3753, 12)
fault.npy	3D NumPy array (binary)	Array of 5000 **faulty** pulses, each with 3753 time steps, and has 12 unique waveforms.	(5000, 3753, 12)
sample_train_data.xlsx*	Excel (readable)	An excel file containing 2D slices (readable data) to view from **normal.npy** and **fault.npy**	NA

* Supplementary file.

The *primary files* include waveform data in the form of NumPy binary arrays. The authors have decided to report the waveform data in a binary format due to their 3D shape (see Eq.1), which makes it extremely difficult and inefficient to report them in excel or CSV formats.

The training data below the "**train**" subdirectory are saved into two NumPy files: **normal.npy** which has the normal pulses, and **fault.npy** which has the faulty pulses. Each file has 5000 pulses. The labels can be implicitly derived from the file name, where the pulses in **normal.npy** have a label of "1″, while the pulses in **fault.npy** have a label of "0″. These training data can be used to build and train the model (statistical, machine learning, etc.). The shape of the data in both **normal.npy** and **fault.npy** can be expressed in a general form:

$$shape = \left(N_{pulses} \times N_{times} \times N_{features}\right) \tag{1}$$

where N_{pulses} is the number of different samples/pulses collected from the system, which is 5000 normal and 5000 faulty pulses, $N_{times} = 3753$ is the number of time steps for each pulse, and $N_{features} = 12$ is the number of features or waveform types recorded for each pulse. The 12 waveforms ($features$) reported in the dataset **in order** are:

1. A+IGBT-I: The current passing through the IGBT switch of phase A+ (unit: A).
2. A+*IGBT-I: The current passing through the IGBT switch of phase A+* (unit: A).
3. B+IGBT-I: The current passing through the IGBT switch of phase B+ (unit: A).
4. B+*IGBT-I: The current passing through the IGBT switch of phase B+* (unit: A).
5. C+IGBT-I: The current passing through the IGBT switch of phase C+ (unit: A).
6. C+*IGBT-I: The current passing through the IGBT switch of phase C+* (unit: A).
7. Mod-I: Modulator current (unit: A).
8. A-Flux: Magnetic flux density for phase A transformer (unit: -).
9. B-Flux: Magnetic flux density for phase B transformer (unit: -).
10. C-Flux: Magnetic flux density for phase C transformer (unit: -).
11. Mod-V: Modulator voltage (unit: kV).
12. CB-V: Cap bank voltage (unit: V).

Certain waveforms (e.g., A-Flux) are scaled by the controller to offer easier manipulation when viewed on the screen by the operator. These waveforms are provided without a unit. However, the reader can expect the natural unit without scaling for the magnetic flux waveforms is Tesla.

Following the training of the model, the test data below the "**test**" subdirectory can be used for fault prognosis/detection analysis. The authors have performed 21 independent experiments described in the next section with the 12 waveforms recorded for each experiment. The waveform data are saved into the NumPy binary data: **test1.npy, test2.npy, …, test21.npy**. These test files are the core of this dataset, which help in testing the trained models in detecting the faults ahead of time. The shape of the data in these 21 test files follows the same shape as

Table 3
Description of the dataset files/directories under the subdirectory "test".

File	Type	Description	*Shape*
test1.npy	3D NumPy array (binary)	Array of recorded pulses of Test 1 that includes 125 pulses.	(125, 3753, 12)
test2.npy	3D NumPy array (binary)	Array of recorded pulses of Test 2 that includes 102 pulses.	(102, 3753, 12)
test3.npy	3D NumPy array (binary)	Array of recorded pulses of Test 3 that includes 89 pulses.	(89, 3753, 12)
test4.npy	3D NumPy array (binary)	Array of recorded pulses of Test 4 that includes 68 pulses.	(68, 3753, 12)
test5.npy	3D NumPy array (binary)	Array of recorded pulses of Test 5 that includes 84 pulses.	(84, 3753, 12)
test6.npy	3D NumPy array (binary)	Array of recorded pulses of Test 6 that includes 92 pulses.	(92, 3753, 12)
test7.npy	3D NumPy array (binary)	Array of recorded pulses of Test 7 that includes 111 pulses.	(111, 3753, 12)
test8.npy	3D NumPy array (binary)	Array of recorded pulses of Test 8 that includes 133 pulses.	(133, 3753, 12)
test9.npy	3D NumPy array (binary)	Array of recorded pulses of Test 9 that includes 125 pulses.	(125, 3753, 12)
test10.npy	3D NumPy array (binary)	Array of recorded pulses of Test 10 that includes 117 pulses.	(117, 3753, 12)
test11.npy	3D NumPy array (binary)	Array of recorded pulses of Test 11 that includes 85 pulses.	(85, 3753, 12)
test12.npy	3D NumPy array (binary)	Array of recorded pulses of Test 12 that includes 100 pulses.	(100, 3753, 12)
test13.npy	3D NumPy array (binary)	Array of recorded pulses of Test 13 that includes 133 pulses.	(133, 3753, 12)
test14.npy	3D NumPy array (binary)	Array of recorded pulses of Test 14 that includes 110 pulses.	(110, 3753, 12)
test15.npy	3D NumPy array (binary)	Array of recorded pulses of Test 15 that includes 102 pulses.	(102, 3753, 12)
test16.npy	3D NumPy array (binary)	Array of recorded pulses of Test 16 that includes 84 pulses.	(84, 3753, 12)
test17.npy	3D NumPy array (binary)	Array of recorded pulses of Test 17 that includes 91 pulses.	(91, 3753, 12)
test18.npy	3D NumPy array (binary)	Array of recorded pulses of Test 18 that includes 101 pulses.	(101, 3753, 12)
test19.npy	3D NumPy array (binary)	Array of recorded pulses of Test 19 that includes 75 pulses.	(75, 3753, 12)
test20.npy	3D NumPy array (binary)	Array of recorded pulses of Test 20 that includes 108 pulses.	(108, 3753, 12)
test21.npy	3D NumPy array (binary)	Array of recorded pulses of Test 21 that includes 77 pulses.	(77, 3753, 12)
tests_labels.csv	CSV file (readable)	2D array of pulse labels for all 21 tests. "**1**" refers to normal pulse, "**0**" refers to faulty pulse	(133**, 21)
sample_test_data.xlsx*	Excel file (readable)	An excel file containing 2D slices (readable data) to view from **test1.npy, test7.npy,** and **test20.npy**	NA
Plots* • test1.png* • test2.png* • . • . • . • test21.png*	Folder	Folder containing 21 png plots of the first and the last pulses of Test 1 – Test 21 for reference	NA

* Supplementary file.
** Number of rows is set to the test with maximum number of pulses (i.e., Test 13 with 133 pulses). Other tests have a smaller number of pulses, e.g. Test 1 has 125 rows/labels, Test 4 has 68 rows/labels.

in Eq. 1 with the only difference is the value of N_{pulses}, as these tests feature different number of collected pulses. The last primary file is **tests_labels.csv,** which contains the labels of the pulses in the 21 tests, with "1″ represents a normal pulse, and "0″ represents a faulty pulse. Typically, for all 21 tests, the first 26 pulses (about 3 min of data collection) are normal (1) while every pulse afterwards is faulty (0) since it involves a system tuning. Since the labels can be presented in a 2D array and readable form, the CSV format is used instead of the binary format. See Table 3 for further descriptions of the test data.

The *supplementary files* include **load_dataset.py,** which we recommend the users to utilize to manipulate and view the data. Given the 3D and binary nature of the data, we report an excerpt of the binary data in 2D slices for the reader to have an impression of the binary data. For example, **sample_train_data.xlsx** reports the A-Flux waveform (out of 12 waveforms) with 3753 time steps for 200 pulses (out of 5000 pulses) in the normal binary file (**normal.npy**) and the fault binary file (**fault.npy**). Similarly, **sample_test_data.xlsx** reports the A-Flux waveform with 3753 time steps for all pulses in **test1.npy, test7.npy, and test20.npy,** where the remaining tests follow the same structure. Lastly, under the "**test**" subdirectory, we provide 21 png plots (**test1.png, test2.png, ..., test21.png**) of the first and the last pulse for all 12 waveforms and for all 21 tests, which help in the sanity check of the test data being used. Supplementary files are marked with * in Table 1, 2, and 3 to differentiate them from the primary files.

4. Experimental Design, Materials and Methods

4.1. Experimental setup

To facilitate reliable data collection and testing while avoiding the need to interpret the normal operation of the SNS, we have configured the HVCM within the RFTF facility to enable continuous data streaming. The RFTF can operate under conditions similar to those of the SNS, rendering the current experimental setup akin to the SNS. While the HVCM is in operation, the three-phase power (A, B, C) at 13.8 kVAC is transformed and converted to $\pm$1300 VDC through transformers and a six-pulse controlled rectifier circuit. The resultant output voltage is subsequently filtered by two capacitors, which store an adequate charge to generate 1.3 ms pulses. This DC voltage is then supplied to three H-bridge circuits, employing IGBT (insulated-gate bipolar transistor) technology, with a nominal switching frequency of 20 kHz. See Fig. 1 (left) for an illustration of the H-bridge schematic.

Fig. 1. H-Bridge circuit of the HVCM in the RFTF (left), high voltage and housing tank (right) connected to a klystron (red cylinder) [4].

Fig. 2. Screenshot of the RFTF controller screen showing the tuning/setting knobs and live waveform diagrams.

The high-power pulses undergo a voltage amplification process through pulse transformers. Within this setup, the leakage inductance of the pulse transformers, coupled with the resonant capacitors in the three phases (C_A, C_B, C_C), creates a resonant circuit characterized by a gain that varies with frequency. These high voltage pulses emerging from the resonant capacitors are merged and rectified by diodes, resulting in output pulses featuring a ripple frequency of 120 kHz. Subsequently, these pulses undergo filtration and are then directed to the cathode of the klystrons, effectively closing the loop of the HVCM circuit.

Within the HVCM, there exists a protective enclosure and a high voltage tank, illustrated in Fig. 1 (right), while the red cylindrical component represents the load associated with the klystron. To measure the magnetic flux within the transformer cores in the three phases (A-Flux, B-Flux, C-Flux), a Rogowski coil is employed with an accuracy of ± 1 %. Additionally, the modulator voltage (Mod-V) and current (Mod-I) are precisely gauged using a North Star High Voltage pulse compensated divider and a Stangenes current transformer, respectively.

To govern the timing of IGBT gating and maintain the signal values within safe limits, the HVCM in the RFTF utilizes a PXI-based controller. This controller also serves the purpose of setting warning and trip thresholds for various signals, digitizing and storing waveforms, and establishing communication via Ethernet with the control system. This controller stands as the primary data acquisition system employed in this research study. Fig. 2 showcases a screenshot from the controller's computer, displaying the waveform plots for the three-phase flux (A-Flux, B-Flux, C-Flux on the right) and the adjustment knobs employed to fine-tune the modulator. Furthermore, Fig. 2 (on the left) provides a visual representation of the modulator voltage (Mod-V). In Section 2.2, we will delve into the details of how the "start pulse" and "flux compensation" settings are manipulated in this work to induce fault-like effects within the recorded waveforms.

Following the data collection from normal and abnormal operating conditions, we plot the pulses/samples for the A-Flux waveform in Fig. 3. These data are reported in the "**train**" subdi-

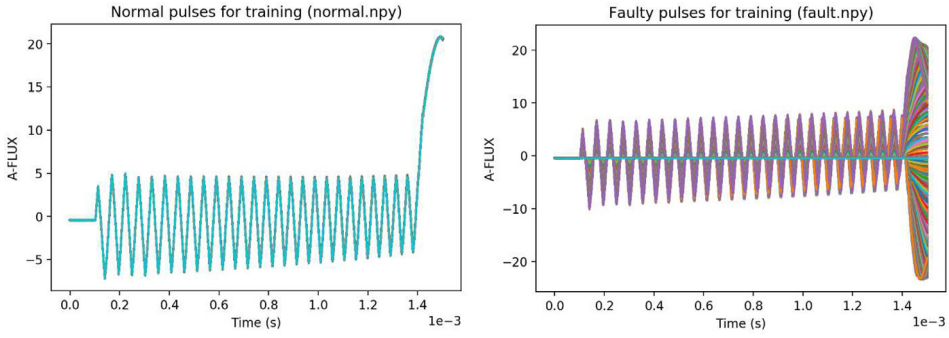

Fig. 3. Plot of the training data: normal A-flux pulses (left) and faulty A-flux (right).

rectory as indicated by Table 2. Given all settings of the RFTF are fixed during normal operation, it is not surprising to see all waveforms in Fig. 3(left) exhibiting identical shape, which is not the case for the faulty waveforms in Fig. 3(right) with different trends and shapes.

4.2. Test collection

The authors have designed 21 experiments with anomalies being induced into the system by systematically changing the settings of the RFTF HVCM, while ensuring the system does not completely fail. The tests mimic the real faults the operators have seen in the past. Basically, each test involves the following steps:

1. The controller and the data acquisition system are set to save a waveform signal every 7 s.
2. The team waits about 3 min at the beginning of each test to collect normal data without any changes in the settings.
3. The team then gradually induces changes in the RFTF settings and waits about a minute to collect waveform data under that change. The settings are changed by **adjusting the 9 knobs** in Fig. 2 to safe values determined by the team. The knobs belong to the categories of "start pulse settings" and "flux compensation settings". Table 4 provides a description of the changes involved in each test.
4. The test finishes when the max/min setting value is reached or when the system is in a dangerous condition to fail. By looking in Table 4, this constraint justifies why the tests have different time lengths.

In general, Table 4 shows that the Tests 1–12 involve a single setting change (e.g., increasing A+ start pulse width, decreasing B-Flux compensation), while Tests 13–21 involve multiple settings being changed simultaneously, which can impose additional risks on the machine. What makes these tests unique is that they include different combinations of the settings in the HVCM that mimic a wide range of fault conditions that HVCM could see in real operation. *The quality of any developed model (machine learning, statistical) is assessed by its ability to predict the first pulse that carries the fault precursors after the 3* min *normal run.*

After the test data collection, we plot sample waveforms from Test 1 and Test 7 in Fig. 4 and Fig. 5, respectively. In each test plot, all 12 waveforms described in Section 1 (A+IGBT-I, A-Flux, Mod-V, etc.) are plotted. In each subplot, the first pulse of the test (before any setting change) and the last pulse (after all settings are changed to their planned values) are shown. This helps the reader to visualize the relative changes in the waveforms from the beginning to the end of the test. These two tests are selected here to reveal how these tests can be tricky. For example, increasing the A+ start pulse width by 5 % every minute in Fig. 4 can make a noticeable variation in most waveforms, which may make it easier for the trained models to

Table 4
Description of the fault prognosis test scenarios [4].

Test num	Description*	Total time (s)
1	A+ start pulse width increases by 5 %/min	868
2	B+ start pulse width increases by 5 %/min	707
3	C+* start pulse width increases by 5 %/min	616
4	A+ start pulse width decreases by 5 %/min	469
5	B+ start pulse width decreases by 5 %/min	581
6	C+* start pulse width decreases by 5 %/min	637
7	A-Flux compensation increases by 25 ns/min	770
8	B-Flux compensation increases by 25 ns/min	924
9	C-Flux compensation increases by 25 ns/min	868
10	A-Flux compensation decreases by 25 ns/min	812
11	B-Flux compensation decreases by 25 ns/min	588
12	C-Flux compensation decreases by 25 ns/min	693
13	A+ start pulse width is set to 20 %, A-Flux compensation increases by 25 ns/min	924
14	A-*/B-*/C- start pulse widths all decrease by 5 %/min	763
15	B+ start pulse width is set to 100 %, B-Flux compensation decreases by 25 ns/min	707
16	A-Flux compensation increases, B-Flux decreases, C-Flux increases by 25 ns/min	581
17	A-Flux compensation decreases, B-Flux increases, C-Flux decreases by 25 ns/min	630
18	C+* start pulse width is set to 90 %, C-Flux compensation increases by 25 ns/min	700
19	B-* start pulse width is set to 50 %, B+ start pulse width increases by 5 %/min	518
20	A+/A-* start pulse widths are set to 20 %, A-Flux compensation decreases by 25 ns/min	749
21	A-*/B-*/C- start pulse widths are set to 40 %, A+/B+/C+* start pulse widths increase by 5 %/min	532

* All tests start with a 3 min of normal run before start inducing fault changes.

predict an upcoming fault. However, Fig. 5 shows that increasing A-Flux compensation by 25 ns every minute only makes a significant variation in the A-Flux waveform, while the other 11 waveforms remain almost identical. This shows that a possible flux fault in the A phase might be missed by the trained models if they rely on the IGBT current signals alone for example. This makes it a challenging task for the analyst to develop a robust model that can handle all 21 tests despite their different nature. The plots of the remaining tests are not reported here for brevity, but they can be found under the "**test**" subdirectory as indicated in Table 3.

4.3. Data usage

Although of the similarity of the general goal of improving HVCM reliability between this dataset and our previous HVCM dataset [5], it is important to clarify major differences between the two datasets:

- This dataset is collected from a test facility well-established to collect large amounts of data (near real-time). The data collected from the RFTF was done in approximately a week of work, without upgrades/changes in the RFTF components, operation settings, or maintenance due to catastrophic failures. Therefore, the uniqueness of this dataset is that it is well-suited for developing models for fault prognosis using the test suite we developed in Table 4. We demonstrated such applications in the main journal article associated with this data article [4].
- Our previous dataset [5] had normal/faulty waveforms from the SNS 15 main HVCMs, which do not have an efficient data acquisition system to handle the streamed data. Therefore, the previous dataset was collected over the course of 2 years with a very sparse time frame

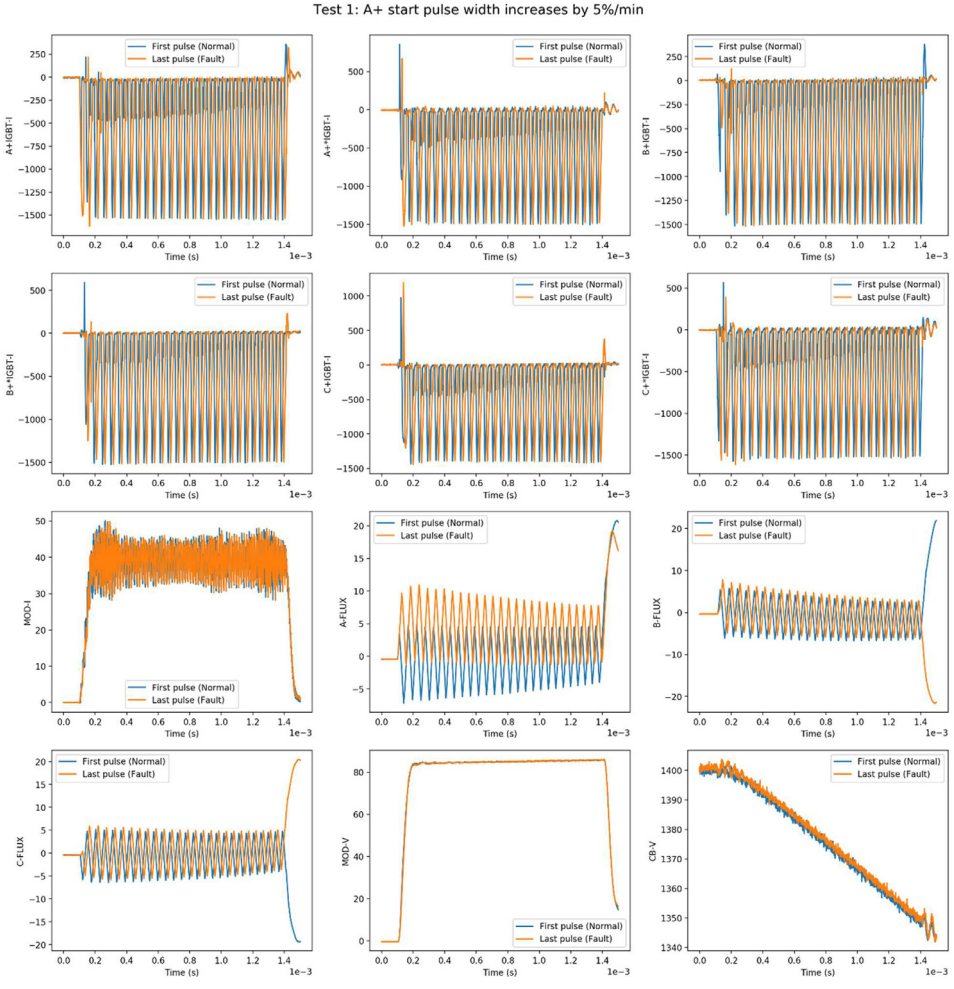

Fig. 4. Waveform results for Test 1. The first pulse in the test (which is before inducing any setting change) and the last pulse (which is after inducing all setting changes) are shown [4].

(recorded waveforms can be separated by hours and even days), where the systems may have experienced changes in its components, operation settings, and routine maintenance. The uniqueness of that dataset [5] is that the controller records large number of faults and the fault reason for different systems, which makes it useful for fault classification in multi-system environments but cannot be used for real-time prediction/prognosis as the waveforms before the fault event were not recorded.

Accordingly, it is worth noticing that the current dataset, which is based on a week of data collection in the RFTF, has a size of 4 GB, while the previous dataset [5] has only 2 GB of data, even though it was collected over 2 years from 15 different HVCM machines in the SNS. These two datasets can complement each other to predict the failure ahead of time and predict the fault reason. Also, upgrading the data acquisition system to allow real time streaming is one of the by-products of this work.

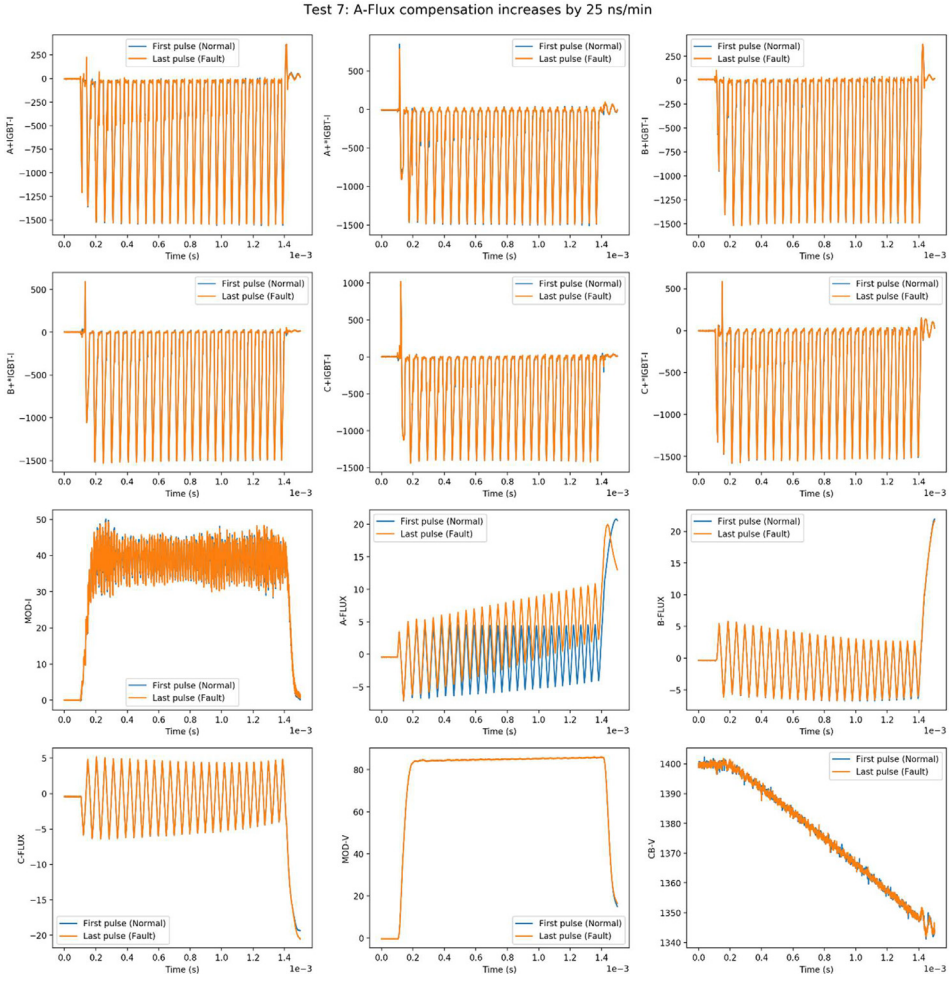

Fig. 5. Waveform results for Test 7. The first pulse in the test (which is before inducing any setting change) and the last pulse (which is after inducing all setting changes) are shown.

Ethics Statements

All methods were carried out in accordance with guidelines and regulations at the Oak Ridge National Laboratory. In addition, this article was approved by the export control office at the Oak Ridge National Laboratory.

Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05–00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Data Availability

Operational Data for Fault Prognosis in Particle Accelerators with Machine Learning (Original data) (Mendeley Data).

CRediT Author Statement

Majdi I. Radaideh: Conceptualization, Methodology, Software, Data curation, Visualization, Formal analysis, Writing – original draft; **Chris Pappas:** Conceptualization, Software, Data curation, Writing – review & editing; **Mark Wezensky:** Software, Data curation, Writing – review & editing; **Sarah Cousineau:** Conceptualization, Funding acquisition, Resources, Writing – review & editing.

Acknowledgments

The authors are grateful for the support from the Neutron Sciences Directorate at ORNL in the investigation of this work. This work was supported by the DOE Office of Science under grant DE-SC0009915 (Office of Basic Energy Sciences, Scientific User Facilities program). A portion of this research used resources at the Spallation Neutron Source, a DOE Office of Science User Facility operated by the Oak Ridge National Laboratory.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Henderson, et al., The spallation neutron source accelerator system design, Nucl. Instrum. Methods. Phys. Res. A 763 (2014) 610–673, doi:10.1016/j.nima.2014.03.067.
[2] D.J. Solley, D.E. Anderson, G.P. Patel, V.V. Peplov, R. Saethre, M.W. Wezensky, HVCM topology enhancements to support a power upgrade required by a second target station (STS) at SNS, in: 2012 IEEE International Power Modulator and High Voltage Conference (IPMHVC), IEEE, 2012, pp. 362–365, doi:10.1109/IPMHVC.2012.6518755.
[3] M.I. Radaideh, C. Pappas, J. Walden, D. Lu, L. Vidyaratne, T. Britton, K. Rajput, M. Schram, S. Cousineau, Time series anomaly detection in power electronics signals with recurrent and ConvLSTM autoencoders, Digit. Signal Process. 130 (2022) 103704, doi:10.1016/j.dsp.2022.103704.
[4] M.I. Radaideh, C. Pappas, M. Wezensky, P. Ramuhalli, S. Cousineau, Early fault detection in particle accelerator power electronics using ensemble learning, Int. J. Prognost. Health Manag. 14 (1) (2023), doi:10.36001/ijphm.2023.v14i1.3419.
[5] M.I. Radaideh, C. Pappas, S. Cousineau, Real electronic signal data from particle accelerator power systems for machine learning anomaly detection, Data Br. 43 (2022) 108473, doi:10.1016/j.dib.2022.108473.