Anomaly Detection Using Machine Learning Techniques for Beam Injections from the SPS to the LHC at CERN

MARC FERRIGGI

Supervised by Dr. Gianluca Valentino

Department of Computer Science
Faculty of ICT
University of Malta

May, 2019

A FYP submitted in partial fulfilment of the requirements for the degree of B.Sc. (Hons.) Computing Science AND Statistics and Operations Research.

Statement of Originality

I, the undersigned, declare that this is my own work unless where otherwise acknowledged and referenced.

Candidate Marc Ferriggi

Signed ____________________

Date April 14, 2019

Acknowledgements

Abstract

Contents

1 Introduction **1**

2 Background and Literature Review **4**

2.1 Beam Instrumentation 4

2.2 Feature Scaling and Reduction Techniques 5

2.3 Unsupervised Anomaly Detection Techniques 6

2.4 Anomaly Detection at CERN 7

2.5 Software Implementation 8

3 Methodology **9**

3.1 Data Collection 9

3.2 Data Cleaning and Analysis 9

3.2.1 TDI BLMs 10

3.2.2 Abort Gap 11

3.2.3 SPS and LHC Intensities 12

3.3 Feature Selection 12

3.4 Merging the Dataset 12

References **13**

List of Figures

1.1 The CERN Particle Accelerator Complex 2
1.2 Anomalous Injections 2

3.1 BLM Histogram 10
3.2 BLM Time Series 10
3.3 BLM Correlation Plot 10
3.4 Abort Gap Histogram 11
3.5 Change in Abort Gap Histogram 11
3.6 Abort Gap Time Series 11

List of Tables

List of Abbreviations

AGM Abort Gap Monitor . . . 5
BCT Beam Current Transformers . . . 5
BLM Beam Loss Monitors . . . 4
BPM Beam Position Monitors . . . 5
CERN European Organization for Nuclear Research . . . 1
DBSCAN Density Based Spatial Clustering of Applications with Noise . . . 6
GMM Gaussian Mixture Model . . . 7
Gy/s Grays per Second . . . 4
IQC Injection Quality Check . . . 2
LHC Large Hadron Collider . . . 1
LOF Local Outlier Factor . . . 6
LS Logging Service . . . 1
MJ Mega Joule . . . 1
MKD horizontally deflecting extraction kicker magnets . . . 5
mm millimetres . . . 5
MSE Mean Square Error . . . 3
PCA Principal Component Analysis . . . 5
RF Radiofrequency . . . 5
SPS Super Proton Synchrotron . . . 1
TDI Beam Absorber for Injection . . . 4
TeV teraelectronvolts . . . 1
TIMBER the user interface to the LS . . . 9
TL Transfer Line . . . 4

1 Introduction

The Large Hadron Collider (LHC) is a "two-ring-superconducting-hadron accelerator and collider" installed at the European Organization for Nuclear Research (CERN) between the years 1984 and 1989 [1]. The collider is 26.7km long and its purpose is to accelerate and collide two proton beams [2].

In order to fill the LHC to its required centre-of-mass energy of 14 teraelectronvolts (TeV), twelve injections from the Super Proton Synchrotron (SPS) consisting of a number of electron bunches of around 1 Mega Joule (MJ) of stored energy are required [3]. Thus, in order to fill the LHC, approximately 4 minutes per beam is required. Furthermore, the whole experiment process of filling the LHC, performing the required checks, running the tests and dumping the beam should take a theoretical minimum of 70 minutes [1]. However, this is expected to take around 6 times longer due to unsuccessful or anomalous proton injections [1].

Clearly, filling the LHC is a challenging task given the high energy of the beam, the very small apertures and the delivery precision's tight tolerances. Thus, multiple censors are installed around the CERN particle accelerator complex [4] which gather readings and data that can be used to check the quality of the injected beam.

For this particular study, data generated from the censors around the injection from the SPS to the LHC will be of particular interest (Figure 1.1). This data is stored using CERN's Logging Service (LS) [5]. While many studies have been made using this logged data and lots of statistical tests have been done with regards to injection quality checks for the LHC (such as [3] and [6]), no literature was uncovered where researchers used unsupervised machine learning methods to analyse this particular data.

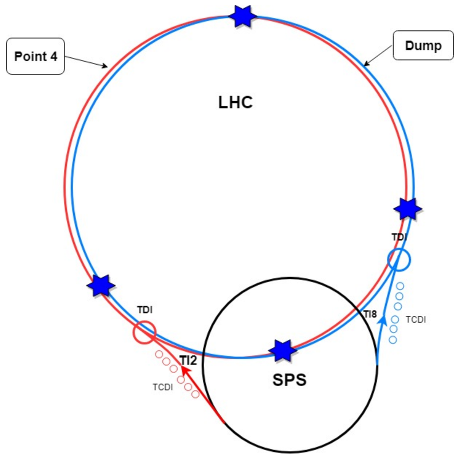

Figure 1.1: Diagram of the particular area of interest of the CERN Particle Accelerator Complex for this study

Furthermore, the Injection Quality Check (IQC) software currently installed has a set of hard-coded rules for detecting anomalies in the SPS-LHC injection [3], however there are documented cases in the past where situations occurred which were outside the originally foreseen rules and were therefore not caught as anomalies. Apart from causing experiments to fail, these anomalous injections could be very costly as a lot of data must be examined after such failures which wastes time that could be used to run more experiments [7]. The major cause of these anomalies is due to the fact that the machine is so large, and needs to be so precise, that minor ground motions over time affect the tilts in the quadrupole magnets which thus affect the orbit of the beam. Figure 1.2 highlights two possible cases of anomalous injections. The first case shows what

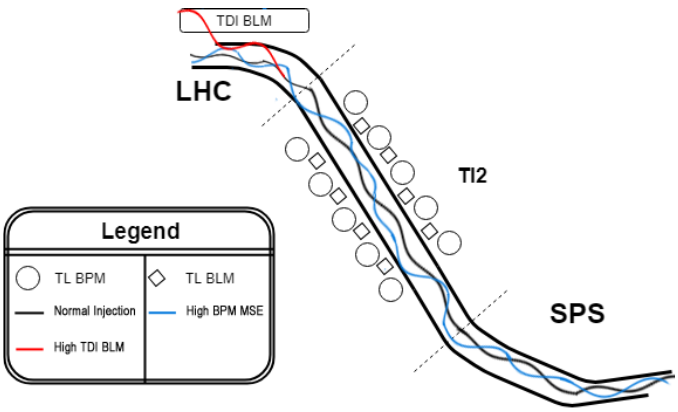

Figure 1.2: Examples of Anomalous Beam Injections

happens to the beam when the BPM gives a high Mean Square Error (MSE) reading. The second case shows what happens to the beam when there is a high loss recorded by the TDI BLM.

The purpose of this study is to apply unsupervised anomaly detection algorithms to try solve the problem of detecting anomalous injections with the hopes of finding a technique that will detect the anomalies not being picked up by the IQC. This can help researchers understand the source of these anomalies and improve the LHC machine availability and performance reach in terms of beam lifetime, beam stability and luminosity.

2

Background and Literature Review

2.1 Beam Instrumentation

Throughout this study, data recorded as the beam leaves the SPS and enters the LHC will be used as input parameters to the chosen anomaly detection algorithms. This data was recorded using different censors located in different parts of the injection life cycle. This section describes the different types of censors that were used to collect the data, highlighting the particular details which need to be considered when analysing this data.

The Beam Loss Monitors (BLM) are some of the most safety critical modules of the LHC because a loss of a very small fraction of this beam may damage parts of the machine or cause a quench in the superconducting magnets [8]. A high beam loss reading could also indicate over-injection. In fact, an injection of a high intensity beam into the LHC is only allowed if there is a low intensity bunch circulating the LHC in order to avoid settings errors [6]. The BLM module is the mostly used module in the current IQC software checks [3]. The BLMs must be reliable; the probability of not detecting a dangerous loss was found to be 5×10^{-6} per channel and they are only expected to generate 20 false dumps per year [8]. The BLMs are extensively logged to a database for offline analysis [8].

For this particular study, the readings logged for the Beam Absorber for Injection (TDI) BLMs and the Transfer Line (TL) BLMs in TI2 and TI8 will be used (refer to Figure 1.1). These readings come in 10 second windows around the injection of a bunch in Grays per Second (Gy/s).

The Beam Position Monitors (BPM) were installed as a system for fast monitoring of the beam's position with respect to it's orbit drift [9]. The trajectory offsets recorded by the BLMs in the transfer lines must be minimised in order to reduce losses [3]. In fact, if the change in orbit substantially exceeds its provided boundary values then the beam should be dumped [9] so as to not cause any damage to the equipment. Unlike the TDI BLMs, the BPM system is independent to the collimator system. For this study, the readings from the transfer line BPMs around TI2 and TI8 will be used (refer to Figure 1.1). Raw values for these readings are stored by the LS in millimetres (mm) and are logged every 1 - 5 seconds on average.

When filling the LHC, it is necessary to keep an abort gap of at least $3\mu s$ in order to accommodate for the horizontally deflecting extraction kicker magnets (MKD) rise time [10]. As the LHC is filling to nominal intensity, this gap will be populated with untrapped particles and particles leaking out of their Radiofrequency (RF) buckets [10]. The Abort Gap Monitor (AGM) was hence specifically designed to measure this particle population in the abort gap [11]. This monitor can be found in Point 4 (refer to Figure 1.1) in the LHC [11]. The raw values extracted for this study are stored in number of particles and come in 10 second groups around the moment of injection.

The actual intensities of the circulating beam are measured by Beam Current Transformers (BCT). For the LHC in particular, a fast BCT is used which is capable of monitoring a broad range of currents as it must be able to detect a single pilot bunch circulating the machine (of 10 μA) as well as the full nominal machine (over 0.5 mA) [12]. These readings are then converted from amps to number of protons per beam and stored for analysis. The intensities for the LHC come in 10 second groups around the moment of injection while the intensities for the SPS give a single value of the intensity at the time of injection.

2.2 Feature Scaling and Reduction Techniques

Feature Scaling and Feature Reduction are two important pre-processing steps that should be considered when using machine learning in the data science process. Standard Scaling in particular will be used in this study as a pre-processing step to Principal Component Analysis (PCA). Standard Scaling ensures that all the features have the properties of a standard normal distribution [13], which is especially important since PCA involves finding the components that maximise the variance [14].

Apart from scaling, another challenge for outlier detection algorithms is data involving high dimensions since the contrast between different points diminishes as the number of dimensions increases [15]. This phenomenon is known as 'The Curse of Dimensionality' and a technique to reduce the effect of this phenomenon is to use a dimension reduction technique and run the outlier detection algorithm on this new lower-dimensioned dataset. In this study, PCA will be used as a dimension reduction technique.

PCA uses statistical and mathematical techniques to reduce the dimension of large data sets, thus allowing a large data set to be interpreted in less variables called principal components [16]. This technique works with the hope that the variance explained by an acceptably small number of principal components is large enough to explain the underlying structure of the dataset reasonably [14]. In fact, this non-parametric method has been used as a means of revealing the simplified structures' underlying complex datasets with minimal effort. The fact that this technique is non-parametric gives it the advantage that each result is unique and only dependent on the provided data set since no parameter tweaking is required [14], however this is also a weakness of PCA as there is no way of exploiting prior expert knowledge on the data set.

2.3 Unsupervised Anomaly Detection Techniques

Unsupervised machine learning algorithms refer to the class of machine learning algorithms where only the input features are available to the learner as there is no access to output labels corresponding to each input feature vector, or the aim of the algorithm is simply to observe or detect patterns in the available data. A. Hyvärinen states in [17] that some of the goals of unsupervised learning include data visualisation, noise reduction, feature extraction and finding interesting components; all of which are of particular interest in this study.

Density Based Spatial Clustering of Applications with Noise (DBSCAN) and Local Outlier Factor (LOF) will both be used as unsupervised anomaly detection algorithms to detect and classify anomalous injections of the past year. Furthermore when working in 3 dimensions or less, these points can also be visualised to help the reader understand better the cause of these anomalies.

DBSCAN was created out of the necessity of having a clustering algorithm with the following requirements:

1. "Minimal requirements of domain knowledge to determine the input parameters,"
2. "Discovery of clusters with arbitrary shape," and
3. "Good efficiency on large databases" [18]

DBSCAN manages to attain these requirements by viewing clusters as "areas of high density separated by areas of low density" [19]. The points with a lower density will thus be considered as anomalies when compared to the regular clusters which have a higher density. This algorithm also introduces the concept of *core samples* which was then used in the design of other unsupervised anomaly detection algorithms such as LOF.

The word *factor* in LOF refers to a "degree of outlier-ness" that this algorithm considers for each point in the data rather than using the concept that "being an outlier is binary" [20]. This algorithm uses a clustering technique which takes concepts from DBSCAN to measure the LOF of each point where a LOF value greater than 1 implies that the point has a lower density than its neighbours and is thus probably an outlier.

2.4 Anomaly Detection at CERN

In the paper released entitled "Opportunities in Machine Learning for Particle Accelerators" [21], it was stated that due to the "large number of process variables, non-linear behaviour, and many interacting subsystems," conventional analysis techniques on today's particle accelerator data is often insufficient and thus machine learning could be used as a means of anomaly detection. Furthermore, the authors also stated that these techniques could be used to detect "subtle behaviours of key variables prior to negative events" and they can also be used to "identify and throw away bad signals."

In his Master's Thesis, A. Halilovic used anomaly detection techniques solely on data obtained from the injection kicker magnets [7]. Halilovic made use of a Gaussian Mixture Model (GMM) and Isolation Forests to detect anomalies however found that the best performance achieved by his proposed pipeline "leaves something to be desired" as too many anomalies were not correctly classified. The author also goes on to suggest that analysing LHC data using the LOF class provided in '*scikit-learn*' could lead to interesting results.

Wielgosz, *et. al.* also wrote a scientific paper on using anomaly detection techniques on the LHC magnets [22]. This time, the authors went for a supervised approach and used Recurrent Neural Networks. They found that using adaptive quantisation to reduce 20-bit inputs into a 4-bit representation was an essential step in improving the algorithm's performance. The authors also stated that these anomaly detection techniques being proposed should not only be considered useful for CERN equipment but also useful in the broader field of anomaly detection on time series data.

In 2017, Valentino *et. al.* released a paper on using anomaly detection techniques "to detect minor changes in the loss maps over time due to collimator settings errors or orbit variations" [2]. The authors used PCA as a dimension reduction technique and then applied LOF on the resulting 2 dimensional data. Their proposed method was shown to positively identify these anomalous loss maps based solely on BPM and BLM readings. Furthermore, they proposed using this technique to monitor losses during fills of the LHC.

2.5 Software Implementation

Although performance of k-means and k-Nearest Neighbours is not as optimal as in other Python packages such as '*PyMVPA*' [23] or '*shogun*' [24] (see Table 1 in [25]), it was decided to use the '*scikit-learn*' machine learning package for this study due to its "state-of-the-art implementation" and "easy-to-use interface tightly integrated with the Python language" [25]. Furthermore, the algorithms implemented using this package can be "used as building blocks for approaches specific to a use case" [25] which will be useful if one would like to extend the scope of this study.

3

Methodology

3.1 Data Collection

The data used in this study was collected from the user interface to the LS (TIMBER) with the help of Dr. Gianluca Valentino who has access to the CERN Intranet. Data was collected from the instrumentation discussed in Section 2.1 and covers 1624 Injections over a time period of 3 months (from 17th August to 20th October 2018). During this time, approximately 65 experiments were performed.

The file sizes for the data gathered from each instrument ranged from 4 KB to 2 MB, these were initially individually analysed (refer to Section 3.2) and then merged to create the dataset used to run the anomaly detection algorithms on (refer to Section 3.4). The total size of the merged datasets were 231 KB and 324 KB for Beam 1 and Beam 2 respectively. Loading this data in memory was not an issue since the file size is rather small, thus the problem of dealing with Big Data was not encountered in this study.

3.2 Data Cleaning and Analysis

After Data Extraction, the provided datasets were analysed separately in order to understand their nature, remove any outliers and be able to aggregate the data correctly for further analysis. In this section the results of this analysis will be presented with the hopes that the reader will have a more clear understanding of later results. Note that all the steps mentioned here were repeated for both beams.

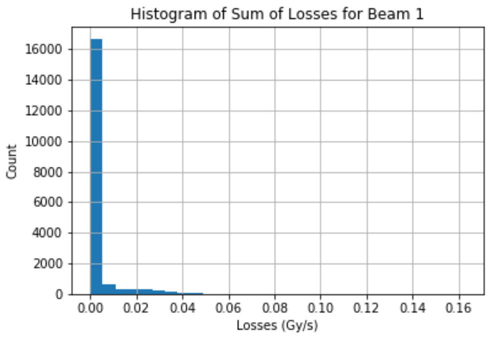

Figure 3.1: Histogram of Sum of Losses for Beam 1

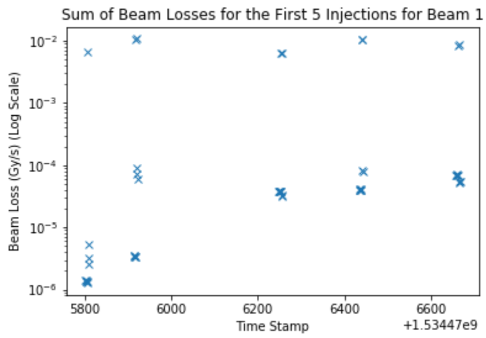

Figure 3.2: Time Series of Beam Loss Sum for the First 5 Injections

3.2.1 TDI BLMs

There are three BLMs in the TDI, each one giving 10 readings around the moment of injection. In order to get a total loss for each injection, the sum of each reading from the 3 monitors was taken (3.1). From the plot of this data (Figure 3.2) it was noted that at the exact moment of injection, there was a spike in the amount of beam lost. Thus, in order to then obtain a single reading corresponding to that particular injection, the maximum sum of losses for each 10 second window was kept.

Once the relevant readings were kept, the sum column was dropped and this data set was saved to be used for anomaly detection. Furthermore, after scaling these points us-

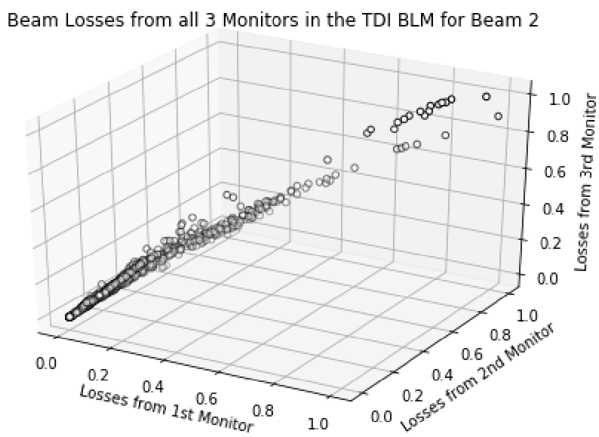

Figure 3.3: Beam Losses from all 3 Monitors in the TDI BLM for Beam 2 after MinMax Scaling

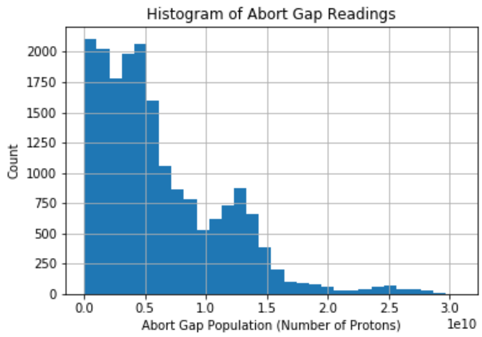

Figure 3.4: Histogram of Abort Gap Population for Beam 1

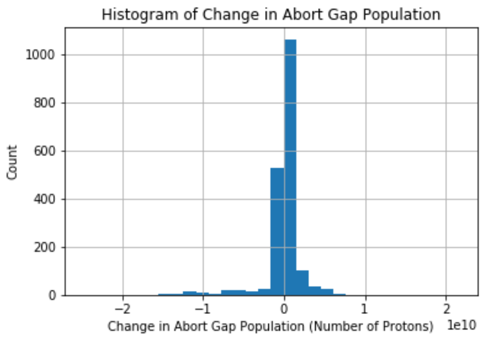

Figure 3.5: Histogram of Change in Abort Gap Population for Beam 1

ing MinMax scaling, it was noted from Figure 3.3 that the readings from the 3 monitors are highly correlated. This was confirmed by computing the correlation matrix which gave a Pearson Correlation value > 0.98 for all pairwise comparisons.

3.2.2 Abort Gap

Similar to the TDI BLM readings, the Abort Gap readings also come in groups of 10 readings around the moment of injection. In this case however, the change in Abort Gap population is of interest for this study, thus the difference between every 10^{th} reading was kept and saved to be used for anomaly detection.

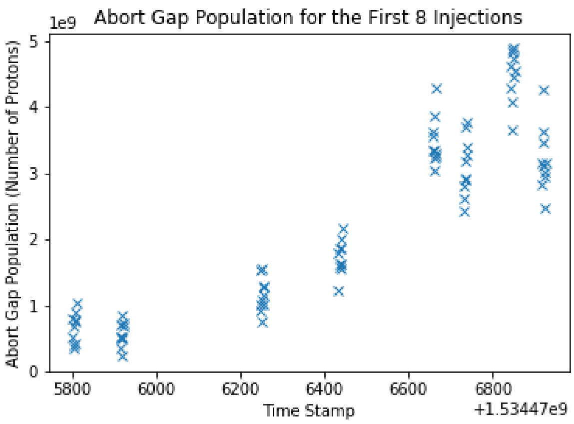

Figure 3.6: Time Series of Abort Gap Population for the First 8 Injections

Figures 3.4 and 3.5 show the histograms of the Abort Gap Population and the Change in Abort Gap Population respectively. A time series plot of the Abort Gap Readings can be seen in Figure 3.6.

3.2.3 SPS and LHC Intensities

3.3 Feature Selection

3.4 Merging the Dataset

References

[1] L. Evans and P. Bryant, “LHC Machine,” *JINST 3*, pp. 1–7, 2008.

[2] G. Valentino *et al.*, “Anomaly Detection for Beam Loss Maps in the Large Hadron Collider,” presented at the 8th Int. Particle Accelerator Conference, 2017.

[3] L. N. Drosdal *et al.*, “Automatic Injection Quality Checks for the LHC,” in *Proc. ICALEPCS*, 2011, pp. 1077–1080.

[4] C. Lefevre, “The CERN Accelerator Complex,” CERN, Tech. Rep., 2008.

[5] C. Roderick, L. Burdzanowski, and G. Kruk, “The CERN Accelerator Logging Service - 10 Years in Operation: A Look at the Past, Present and Future,” presented at the 14th Int. Conf. Accelerator & Large Experimental Physics Control Systems, 2013.

[6] V. Kain *et al.*, “Injection Beam Loss and Beam Quality Checks for the LHC,” in *Proc. of IPAC*, 2010, pp. 1671–1673.

[7] A. Halilovic, “Anomaly Detection for the CERN Large Hadron Collider Injection Magnets,” Master’s thesis, KU Leuven, 2018.

[8] E. Holzer *et al.*, “Beam Loss Monitoring System for the LHC,” presented at the IEEE NSS, 2006.

[9] *Protection of the CERN Large Hadron Collider*, ser. New Journal of Physics, vol. 8, no. 290, CERN, 2006. [Online]. Available: http://www.njp.org/

[10] M. Meddahi *et al.*, “LHC Abort Gap Monitoring and Cleaning,” presented at the IPAC, 2010.

[11] T. Lefevre *et al.*, “First Operation of the Abort Gap Monitors for LHC,” CERN, Tech. Rep., 2010.

[12] O. R. Jones, “LHC Beam Instrumentation,” in *Proc. of PAC*, 2007, pp. 2630–2634.

[13] (2019, April) Importance of Feature Scaling. [Online]. Available: scikit-learn.org

[14] J. Shlens, “A Tutorial on Principal Component Analysis,” April 2014.

[15] A. Zimek, E. Schubert, and H.-P. Kriegel, "A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data," *Statistical Analysis and Data Mining 5*, 2012.

[16] M. Richardson, "Principal Component Analysis," May 2009, Class Lecture.

[17] A. Hyvärinen, "Unsupervised Machine Learning," Lecture Notes, University of Helsinki.

[18] M. Ester *et al.*, "A Density Based Algorithm for Discovering Clusters," in *Proc. KDD-96*, 1996, pp. 226–231.

[19] (2018, November) Clustering. [Online]. Available: scikit-learn.org

[20] M. Breunig, H. Kriegel, and J. Sander, "LOF: Identifying Density-Based Local Outliers," in *Proc. ACM SIGMOD 2000 Int. Conf. On Management of Data*, 2000.

[21] A. Edelen, C. Mayes, and D. Bowring, "Opportunities in Machine Learning for Particle Accelerators," *ArXiv*, 2018.

[22] M. Wielgosz *et al.*, "The Model of an Anomaly Detector for HiLumi LHC magnets based on Recurrent Neural Networks and Adaptive Quantization," *ArXiv*, 2017.

[23] (2018, November) PyMVPA Developer Guidelines. [Online]. Available: pymvpa.org

[24] (2018, November) The Shogun Machine Learning Toolbox. [Online]. Available: pypi.org/project/shogun-ml

[25] *Scikit-learn: Machine Learning in Python*, ser. Journal of Machine Learning Research, vol. 12, 2011.