ISBN: 978-3-95450-210-3 doi:10.18429/JACoW-FEL2019-WEB03

APPLICATION OF MACHINE LEARNING TO BEAM DIAGNOSTICS

E. Fol* [1 2], R. Tomás[1], G. Franchetti[2 3], J. Coello de Portugal[4], [1]CERN, 1211 Geneva 23, Switzerland
[2]Johann-Wolfgang Goethe University, 60438 Frankfurt am Main, Germany
[3] GSI Helmholtzzentrum für Schwerionenforschung, 64291, Darmstadt, Germany
[4]Paul Scherrer Institut, 5232, Villigen, Switzerland

Abstract

Machine Learning (ML) techniques are widely used in science and industry to discover relevant information and make predictions from data. The application ranges from face recognition to High Energy Physics experiments. Recently, the application of ML has grown also in accelerator physics and in particular in the domain of diagnostics and control. The target of this paper is to provide an overview of ML techniques and to indicate beam diagnostics tasks where ML based solutions can be efficiently applied to complement or potentially surpass existing methods. Besides, a short summary of recent works will be given demonstrating the great interest for use of ML concepts in beam diagnostics and latest results of incorporating these concepts into accelerator problems, with the focus on beam optics related applications.

MOTIVATION

Traditional optimization tools demonstrate successful performance in applications on linear optics corrections and problems with limited amount of optimization targets [1–6]. Bigger challenges emerge when diagnostics of complex non-linear behavior is required and several variables have to be taken into account as final objective. The amount of time and computational power required by traditional methods might become unacceptable for future accelerator facilities.

ML is well known for surpassing human performance in some specific tasks such fraud detection, forecasting of market trends and risks, online recommendations, recognition of voice and images and in general in discovering correlations in large scale data sets. Most of these tasks can find analogies in beam control and diagnostics. For example, anomaly detection methods applied for fraud detection can be used to detect defects in the instrumentation and forecasting techniques can be transferred to predict beam behavior during operation.

Free Electron Lasers (FEL) problems for optimization and diagnostics have to deal with non-linear, multi-objective functions which depend on thousands of time-varying machine components and settings. These properties meet the limitations of traditional optimization methods and make this problem a perfect candidate for application of ML-based techniques. The main limitation of traditional optimization methods is that the objective function or specific rules and thresholds have to be known. ML methods can learn from given examples without requiring explicit rules.

* elena.fol@cern.ch

RELEVANT MACHINE LEARNING CONCEPTS

ML techniques aim to build computer programs and algorithms that automatically improve with experience by learning from examples with respect to some class of task and performance measure, without being explicitly programmed [7]. Depending on the problem and existence of learning examples, different approaches are preferred. If pairs of input and desired output are available, an algorithm can generalize the problem from the given examples and produce prediction for unknown input. ML algorithms that learn from input/output pairs are called *supervised learning* algorithms. Opposite to supervised learning, *unsupervised learning* algorithms solve the tasks where only input data is available. Unsupervised learning is suitable for the problems such anomaly detection, signal denoising, pattern recognition, dimensionality reduction and feature extraction. In the following a brief overview on significant machine learning concepts that can be used as supervised as well as unsupervised approaches is presented. We also give a short introduction to Reinforcement Learning - ML technique which recently became of great interest especially for control tasks.

Artificial Neural Network

Artificial Neural Networks (ANNs) are well suited for learning tasks, where data is represented by noisy, complex signals and the target output function may consist of several parameters. A basic ANN consists of a single processing unit (*neuron*), that takes the *weighted* inputs and an additional activation function to introduce the nonlinearity in the output. For more complex practical problems, ANNs are composed of several interconnected *hidden layers* with multiple neurons stacked. ANNs can be used for both regression and classification problems. In case of classification the output can be either a class label or a probability of an item belonging to a class. The learning of ANN is performed using *backpropagation* algorithm [8] on a set of examples. For each example the training algorithm computes the derivatives of the output function of the network. The obtained gradients with respect to all weights are then used to adjust the weights in order to achieve a better fit to the target output. In backpropagation *stochastic gradient descent* or one of its improved extensions [9,10] is applied as optimization method in order to minimize the loss between the network output values and the target values for these outputs by updating the connection weights. ANNs with many hidden layers called *deep neural networks* are able to use fewer neurons per layer and have a better generalization

ISBN: 978-3-95450-210-3 doi:10.18429/JACoW-FEL2019-WEB03

ability [11], however the optimization of the structure and training of these networks is not trivial. There are no strict rules for building ANN architecture (number of neurons, layers, initial weights) as it usually heavily depends on a particular problem. However, techniques to adjust the architecture parameters exist. A detailed overview on various ANN architectures and training methods and their suitability for different applications can be found in [12–14].

Decision Trees and Ensemble Methods

Decision tree learning is a method for approximating discrete-valued target functions, which are represented by decision trees. Considering the case of classification, decision trees sort down the input instances from the root to leaf nodes. Usually, the splitting is based on one of the input parameters or a specified set of splitting criteria [15, 16]. Each leaf corresponds to one class representing the most appropriate class label. For regression problems the leaf nodes correspond to an approximation of target values.

Using a single tree, a model might not be able to generalize and perform poorly on unexplored sample. One possible solution to overcome this problem is to build ensembles of trees [17]. By training several slightly different models and taking the average prediction, the variance of the model can be reduced.

Compared to ANNs, decision trees are simpler to interpret and to understand its way of obtaining the final results and the underlying process, e.g through the feature importance analysis. Feature importance analysis helps to understand the contribution of each input parameter to the decision during the training process. The ability of decision trees to evaluate the importance of input parameter is a significant advantage of these algorithms. Knowing the importance of the features we can reduce the model complexity and simplify the data preprocessing steps without significant accuracy loss.

Clustering

Cluster analysis includes methods of grouping or separating data objects into clusters, such that dissimilarity between the objects within each cluster is smaller than between the objects assigned to different clusters [18, 19]. Cluster analysis is used in a wide range of applications. Data clusters can be considered as a summarized representation of the data, such that group labels can describe patterns or similarities and differences in the data. Moreover, clustering can be used for prediction, such that classification of unseen data is performed based on knowledge about the properties of present data and by evaluating their similarity to the incoming data sample. The significant benefit of cluster analysis is the *unsupervised learning* approach, which means that no labeled data is needed to find a solution.

The simplest and the most commonly used clustering algorithm is k-means [20], which is based on centroid search. Another kind of clustering algorithms are the density-based algorithms such DBSCAN [21], that views clusters as areas of high density separated by areas of low density, instead of looking for the centroids. Decision tree based methods also can be applied for cluster analysis using the data splits based on different features. Most of cluster analysis techniques allow to build clusters in a multidimensional space.

Apart from classification and pattern recognition, cluster analysis can be used as denoising method looking for abnormalities in the signal. Moreover, building clusters combining a large set of different observables can simplify the data visualization and manual analysis, such elimination of outliers in the measurements and detection of anomalies.

Reinforcement Learning

The concept of Reinforcement Learning (RL) is based on environment-agent interaction [22]. The agent takes an action on the environment, and the environment reacts producing a reward, which is used by the agent to learn how to improve its actions. The approach does not require an existing data set consisting of input-output pairs, instead the agent is learning based on the continuous interaction with the environment which is varying depending on the action and its own dynamics. Considering this learning principle, RL can be applied to unstable, time-varying problems since the agent should be able to adjust its action to the changes of the response from the environment. The ability of RL techniques to be applied on time-varying unknown dynamics makes this approach particularly appealing for the control and optimization of accelerator components. Recent advances on RL application on accelerator control tasks can be found in [23].

OVERVIEW ON CURRENT APPLICATIONS

In the following we demonstrate some ML applications currently being used in accelerator technology and ongoing research on potential ML based approaches. An earlier overview on previous works related to beam diagnostics can be found in [24], for a wider overview on opportunities in ML for particle accelerators see [25–27].

Virtual Diagnostics

Various instruments and diagnostics techniques are required in order to monitor the beam itself and variables which affect its parameters. Virtual diagnostics can assist in case a direct measurement would have a destructive impact on the operation or in the locations where no physical instrumentation can be placed. ML can provide techniques to build such virtual beam diagnostics instruments. Simulation studies and experimental demonstrations have been carried out on FACET-II and Linac Coherent Light Source (LCLS) to study ML-based longitudinal phase space (LPS) prediction. Training data for a feed-forward ANN has been acquired from a large number of simulations that represent changes in LPS distribution as response to the change of various accelerator parameters, as well as from the existing measurements at LCLS. ML model demonstrates a good agreement between the prediction and simulated or measured LSP images [28]. Another example is the estimation of

ISBN: 978-3-95450-210-3 doi:10.18429/JACoW-FEL2019-WEB03

oscillation amplitude and synchrotron damping time based on LPS measurements at Shanghai Synchrotron Radiation Facility (SSRF) [29]. Here, Gradient descent algorithm is used to estimate the fitting parameters which are then used as target variables in a supervised model. ANN is trained to predict these values from longitudinal phase measurements obtained from the Beam Position Monitors (BPM). Another example from SSRF is a study on correlations between the beam size and the images from multi-slit imaging system, aiming to improve the accuracy of BPMs using ANN [30].

A special kind of ANN, *convolutional neural networks* (CNN) [31] have been applied at FAST on image based diagnostic during beam operation [32]. A combination of a CNN and a feed-forward NN yields promising results for the prediction of beam parameters on simulated data sets. The model uses simulated cathode images, solenoid strengths and the gun phase as inputs and produces a prediction for various downstream beam parameters. Application of ANN can be found also in correction of distorted beam profiled measured at ionization profile monitors (IPM) [33]. ANN model has been trained on IPM simulations in order to establish the mapping between measured profiles together with bunch length and bunch intensity to the original beam profile.

Optimization and Operation

ML methods are especially suited for non-linear and time-varying systems with large parameter spaces. Operation of a complex system such as an accelerator, whose beam dynamics exhibits nonlinear response to machine settings can be considered as a typical ML task. Due to the constant increase of machine design complexity and development of new interacting systems, traditional techniques might become insufficient. Reinforcement learning demonstrates a great ability to solve complex control tasks [23]. Recently, its application has been studied on control tasks in the domain of accelerators, e.g. for the control of the micro-bunching instability at the KIT storage ring Karlsruhe Research Accelerator (KARA) [34].

ANN based application has been successfully applied at the LCLS to predict x-ray pulse properties by decoding complex hidden correlations between parameters obtained from slow diagnostics such as photon energy and properties measured by fast diagnostics [35]. Another example is the application of intelligent control techniques to maximize the average pulse energy in FELs. The developed techniques allow to tune up to 105 components simultaneously based only on noisy average bunch energy measurements [36].

Beam Optics Correction

Attempts to build beam diagnostics and beam control systems using ML have been made already in the past decades [37–39]. Despite the early stage of ANN technology at the time, the obtained results have shown the potential of supervised learning solution to be applied in beam control tasks, mainly for linear orbit correction.

Apart from supervised learning, optics correction can be approached from a probabilistic point of view as it was recently shown in [40]. In this example, the quadrupole error distribution is fitted using Bayesian approach. Degeneracy of error sources is solved by selecting non-correlated BPM signals.

Supervised learning is being under study with two different approaches for optics correction at the LHC aiming to reduce optics errors by finding quadrupolar gradient errors. In order to compute the corrections, measured data have to be compared with the ideal optics design. The deviations from ideal optics introduced by quadrupolar gradient errors have to be compensated by applying corrections [41, 42]. In terms of machine learning, this task can be defined as a regression problem.

In the first approach, simulations of randomly generated errors in the quadrupoles powered in series (circuits) are used as target values and the optics perturbation produced by these errors is the input of the regression model. To correct the perturbed optics, the circuit errors predicted by the trained regression model just have to be applied with the opposite sign. However, under realistic conditions the errors of every single magnet instead of circuits perturb the optics, which has a different effect compared to the strength change in the circuits. In case the objective is to obtain the circuits settings to be implemented in the LHC, the optics functions in the training data have to be perturbed by errors in circuits in order to build input-output pairs required for supervised learning approach. As input we use phase advance, beta and normalized dispersion deviations from the ideal model simulated as measurements at the BPMs, 2560 features in total. The output variables are the values for the strength change in the circuits (193 target variables). The simulated phase advance measurements are given Gaussian noise relative to the $\sqrt{\beta}$ - function at the location of the BPM. In order to assert the ability of the model to correct the optics under realistic conditions, an additional data set is generated for the validation where the optics is perturbed by single quadrupoles instead of circuits. As figures of merit we use β-beating and the deviation of the normalized dispersion to the ideal model after applying the obtained correction values for the circuit strength. The comparison between different algorithms [43], shows that all applied models perform equivalently, therefore we chose Linear Regression implementation of *Scikit-learn* [44] for further studies since the obtained model is easier to interpret and the training can be performed significantly faster. The advantage of regression model against currently used response matrix approach [45] is the ability of extracting an average linear response over the training population instead of only using the unperturbed model and the response of a single observable to a strength change in a single corrector. Despite the fact that the errors of prediction on training and test set are acceptably small, the predicted corrections are less effective on the validation data which is perturbed with individual quadrupole errors. Nevertheless, the ability of such a model to reduce the optics perturbation to the level comparable with traditional response matrix approach is clearly demonstrated and shows the potential of ML-based optics correction.

ISBN: 978-3-95450-210-3 doi:10.18429/JACoW-FEL2019-WEB03

In the second approach we studied the ability of regression models to address directly the single quadrupole errors. Another simulation dataset is generated in order to train a model to predict the errors of each quadrupole from the phase, beta and dispersion perturbations. The model achieves acceptably high score for R^2 coefficient which is defined as follows:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

where $\hat{y}_i$ is the predicted value of the i-th sample, y is the corresponding true value for n total samples and $\bar{y}$ is the mean of true values. The scores of the model are 0.98 and 0.86 for training and test data set respectively, it has to be noted that the resulting difference between training and test scores exhibits slight overfitting of the model. This issue also explains the high relative error between true and predicted values for the validation set (~20% ± 0.23). Identification of magnet errors in a circular machine is known to be a degenerate problem with multiple solutions. Despite this limitation, the attempt to correct the optics using the ML-model prediction yields impressive results. Figure 1 shows correction results obtained with iterative response matrix approach [46] and ML model using linear regression for 120 LHC simulations. The great correction results achieved with ML model despite the relatively poor performance of the model on the training and test sets can be explained with the fact, that in order to correct the optics, it is sufficient to find one of the multiple solutions which can compensate the introduced optics perturbations fitted into the model.

High R^2 score shows that the model can explain the variance of target values based on all features, but not all of the features are significant to obtain the correct output. Reducing the number of the features by selecting the least correlated BPMs should improve the accuracy of model prediction, prevent overfitting, as well as provide help to deal with the degeneracy. Dimensionality reduction techniques already demonstrated their potential for orbit correction, as well as to be applied to dynamic aperture optimization [47]. Therefore, the next steps of the study is the application of dimensionality reduction techniques, followed by the introduction of non-linear error sources into the data sets and generalization to different optics settings.

Instrumentation Fault Detection

Anomaly detection techniques are suitable for the detection of unusual events that do not conform to expected patterns. They also can be used to identify outliers and remove noise. Anomaly detection can be performed using classification on labeled data (supervised learning), unsupervised learning techniques including clustering or semi-supervised learning methods such as autoencoder, a special ANN representing the model trained on normal data set and then detect the anomalies based on the value of the loss function generated by the representative model on the given test sample [48]. An early example on anomaly detection in beam diagnostics in storage ring is the application of ANN to predict the orbit at particular beam position monitor (BPM)

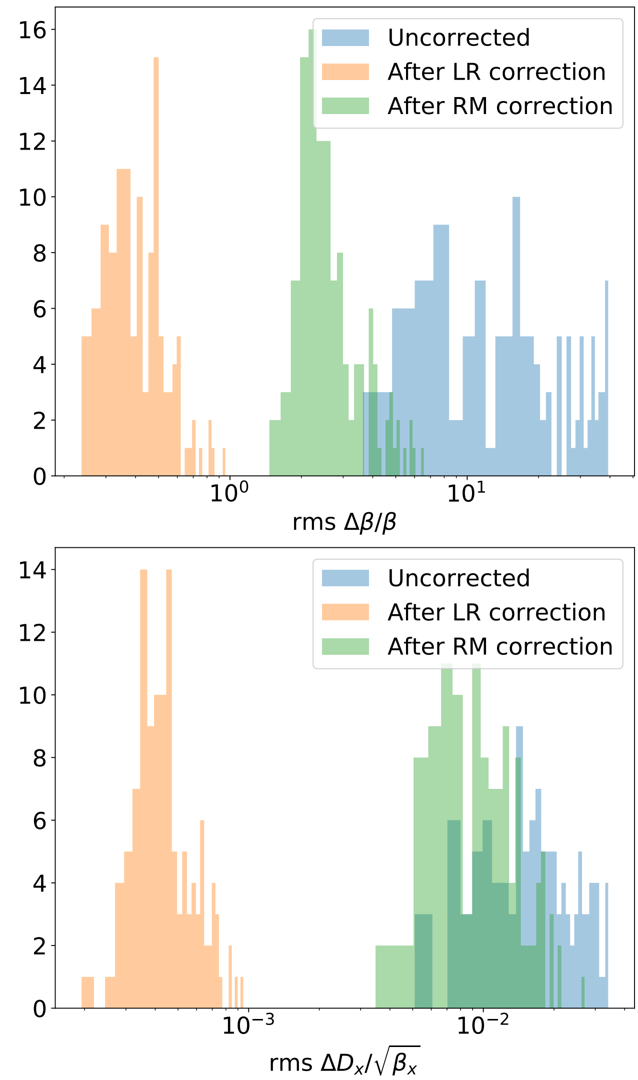

Figure 1: Results on β-beating and normalized dispersion deviation from ideal optics after applying linear regression prediction of individual quadrupole errors (LR) and corrections values for the circuits computed by response matrix (RM). The figure shows rms distribution of 120 simulations.

based on measurements at other BPMs at the Pohang Light Source [49]. A large deviation between measured and predicted orbit should mark malfunctioning BPM.

An example for anomaly detection using unsupervised learning is the detection of faulty BPMs at the LHC [50, 51]. This method recently became a standard part of optics measurements at LHC and has been successfully used during beam commissioning and machine developments for different optics settings in 2018. BPMs measure the beam position at several turns around the machine. The optics functions are then calculated from the harmonic analysis of the turn-by-turn BPM readings. Most of the noise and faulty signals can be removed using predefined thresholds, as well as through applying advanced signal-improvement techniques based on SVD [52] to reduce noise in BPM readings. However few nonphysical values are usually observed

ISBN: 978-3-95450-210-3 doi:10.18429/JACoW-FEL2019-WEB03

in the optics computed from the data cleaned with these techniques. These spikes have to be removed by manually identifying the faulty BPMs, removing them from the harmonic analysis data and repeating the optics analysis, which requires human intervention and loss of valuable machine development time.

Further issue is that the spike does not necessarily appear directly at the location of the faulty BPM, due to the method applied for the optics computation at the LHC [53, 54], so identification of actual BPM faults is not trivial. Moreover, not all reasons for the appearance of BPM anomalies are known, therefore we cannot define thresholds which would indicate remaining faulty BPMs. Giving these constraints, unsupervised learning appears as appropriate technique to detect faulty BPMs prior to optics computations. The appearance of outliers is challenging for the application of centroid or distance based clustering methods. Instead, density-based clustering methods such as DBSCAN and LOF [55] have been applied, however Isolation Forest (IF) algorithm which is a decision tree - based method [56] achieves the best results. Figure 2 shows a comparison between the beta-beating reconstructed from the measurements before and after applying IF. It can be clearly observed that most of the remaining outliers have been removed.

Since the knowledge about actual defective BPMs is not available, the assessment of cleaning algorithms has to be performed on simulations where the actual bad BPMs are known and can be used as labeled data to evaluate the performance of the method. The detailed description of this simulations study can be found in [51]. The main challenge of applying any cleaning methods on the measurements data is, that depending on the chosen algorithm parameter, some of the good BPMs can be wrongly identified as faulty. The comparison of different anomaly detection methods applied on the simulated BPM faults is demonstrated in Fig. 3. In case of large machines such as the LHC equipped with hundreds of BPMs, it is important to decrease the number of faulty signal artifacts as much as possible, because a single faulty BPM can affects the optics computation at multiple locations. The absence of few good BPMs that might be caused by IF algorithm does not have a significant negative effect since the optics computation can be propagated to the next available BPM. Considering smaller machines, it is crucial to keep as much BPM information as possible, removing only critically erroneous signal. In this case, a method such DBSCAN appears to be more appropriate since, as it was shown on simulations [51], the method does not identify any good BPMs as faulty, however a portion of bad BPMs is still remaining in the measurement.

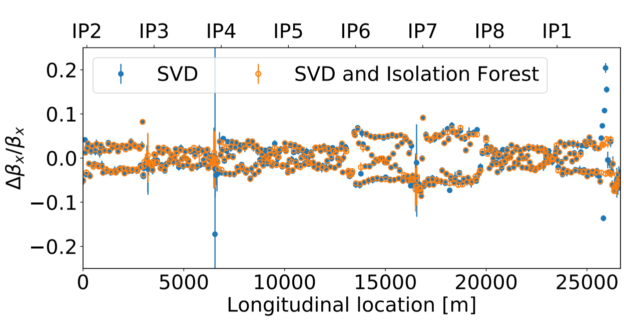

Figure 2: The comparison between beta-beating computed before and after IF cleaning demonstrates that IF anomaly detection significantly reduces the number of nonphysical spikes. The optics is computed for Beam 2 in horizontal plane with β^*=50 cm.

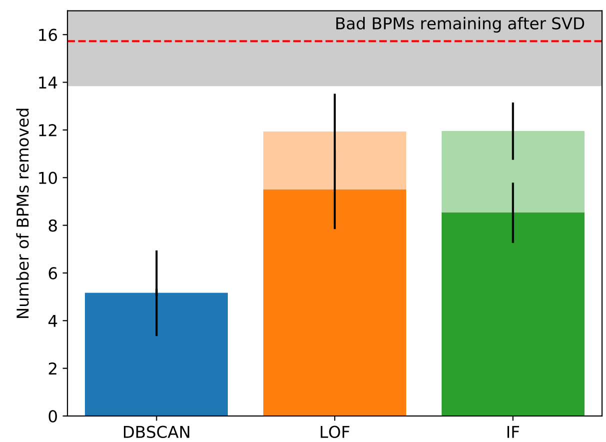

Figure 3: The comparison is carried out on 20 simulations for each plane, the results are averaged. Each bar represents the number of BPMs removed by the method. Dark fraction corresponds to the number of removed BPMs that are actually bad.

CONCLUSION

Typical characteristic of supervised ML tasks is the ability to deal with large amount of structured data. This leads to the conclusion that the implementation of supervised ML solutions requires large existing training data sets or development of appropriate data acquisition tools in order to provide the data in "machine-understandable" format, which is not necessarily available out-of-the-box since the traditional control systems usually imply human intervention. The effort that has to be put on automation such as building data acquisition infrastructure and training of complex models might be more costly and resources expensive than traditional methods. On the other hand, automation of some particular systems using ML as it was done for example, in collimators alignment at LHC [57] is very effective and can save operational resources.

The ability of unsupervised learning to discover unknown patterns in the data is useful especially for anomaly detection tasks such as detection of instrumentation defects, e.g. using clustering for faulty BPMs signal. Such methods can be performed directly without training in arbitrary accelerator systems. The optics correction results achieved with supervised learning convincingly demonstrate the great potential of this approach opening new opportunities for optics control in current and future accelerators.

ISBN: 978-3-95450-210-3 doi:10.18429/JACoW-FEL2019-WEB03

REFERENCES

[1] J. W. Flanagan *et al.*, "A simple real-time beam tuning program for the KEKB injector Linac", in Proceedings of the 1998 International Computational Accelerator Physics Conference, Monterey, CA, KEK Report No. 98-208, 1999.

[2] N. J. Walker, J. Irwin, and M. Woodley, Report No. SLAC-PUB-6207, 1993.

[3] Y. Funakoshi *et al.*, "Performance of KEKB with crab cavities", in Proceedings of the 11th European Particle Accelerator Conference, Genoa, 2008, edited by I. Andrian and C. Petit-Jean-Genaz (Ref. [111]), pp. 1893–1895.

[4] W. Fischer, J. Beebe-Wang, Y. Luo, S. Nemesure, L. K. Rajulapati, "RHIC proton beam lifetime increase with 10- and 12-pole correctors", in Proceedings of IPAC'10, Kyoto, Japan, pp. 4752–4754.

[5] M. Aiba, M. Boege, N. Milas, A. Streun, "Ultra low vertical emittance at SLS through systematic and random optimization", Nucl. Instrum. Methods Phys. Res., Sect. A, 694, pp.133-139, 2012.

[6] X. Huang, J. Corbett, J. Safranek, J. Wu, "An algorithm for online optimization of accelerators", Nucl. Instrum. Methods Phys. Res., Sect. A 726, 77 (2013).

[7] T. Mitchell, "Machine Learning", 1st ed., McGraw-Hill, Inc., New York, USA (1997).

[8] D. Rumelhart, G. Hinton, R. Williams, "Learning representations by back-propagating errors", Nature 323 (6088), 533–536, (1986) .

[9] D. Kingma and J. Ba, "Adam: A method for stochastic optimization", CoRR abs/1412.6980 (2014).

[10] J. Duchi, E. Hazan, Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization", Journal of Machine Learning Research 12 (2010): 2121-2159.

[11] Y. Bengio, Y. LeCun, G. Hinton, "Deep Learning", Nature. 521 (7553): 436–444, 2015.

[12] S. Haykin, "Neural Networks: A Comprehensive Foundation", 2nd ed., Prentice Hall PTR Upper Saddle River, NJ, USA (1998).

[13] J. Schmidhuber, "Deep learning in neural networks: An overview", Neural networks: the official journal of the International Neural Network Society 61 (2015): 85-117.

[14] F.L. Lewis, "Neural Networks in Feedback Control Systems", Mechanical Engineer's Handbook, John Wiley, New York, USA (2005).

[15] O. Maimon and L. Rokach, "Data Mining and Knowledge Discovery Handbook", Springer-Verlag New York, Inc., Secaucus, NJ, USA (2005).

[16] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, "Classification and Regression Trees", Statistics/Probability Series, Wadsworth Publishing Company, Belmont, USA (1984).

[17] T.G. Dietterich, "Ensemble Methods in Machine Learning", In Proceedings of the First International Workshop on Multiple Classifier Systems, Springer-Verlag, London, UK (2000).

[18] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning", Springer Series in Statistics, Springer New York Inc., New York, USA (2001).

[19] A. K. Jain, M. N. Murty, and P. J Flynn, "Data clustering: a review", ACM computing surveys (CSUR), 31(3) (1999).

[20] S. P. Lloyd, "Least squares quantization in PCM", IEEE Transactions on Information Theory, 28 (1982).

[21] M. Ester, H. Kriegel, X. Sander, J.Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96. AAAI Press (1996).

[22] R.S. Sutton, A, G. Barto, "Reinforcement Learning: An Introduction", MA, USA: MIT Press, 2018.

[23] T. P. Lillicrap *et al.*,"Continuous control with deep reinforcement learning", arXiv:1509.02971 (2015).

[24] E. Fol, J. M. Coello de Portugal, and R. Tomás, "Application of Machine Learning to Beam Diagnostics", in *Proc. 7th Int. Beam Instrumentation Conf. (IBIC'18)*, Shanghai, China, Sep. 2018, pp. 169–176. doi:10.18429/JACoW-IBIC2018-TUOA02

[25] A. L. Edelen, S. G. Biedron, B. E. Chase, D. Edstrom, S. V. Milton, P. Stabile, "Neural Networks for Modeling and Control of Particle Accelerators", IEEE Transaction of Nuclear Science, 63 (2), 2016.

[26] A. Edelen *et al.*, "Opportunities in Machine Learning for Particle Accelerators", arXiv:1811.03172 (2018).

[27] S. Biedron, "Adding Data Science and More Intelligence to Our Accelerator Toolbox", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 1191–1197, doi:10.18429/JACoW-IPAC2019-TUZPLM1

[28] C. Emma et.al "Machine learning-based longitudinal phase space prediction of particle accelerators", Phys. Rev. Accel. Beams (21), 112802 (2018).

[29] X.Y. Xu, Y.B. Leng, and Y.M. Zhou, "Machine Learning Application in Bunch Longitudinal Phase Measurement", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 2625–2628, doi:10.18429/JACoW-IPAC2019-WEPGW064

[30] B. Gao, Y.B. Leng, and X.Y. Xu, "Deep Learning Applied for Multi-Slit Imaging Based Beam Size Monitor" in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 2587–2590, doi:10.18429/JACoW-IPAC2019-WEPGW049

[31] Y. LeCun, K. Kavukvuoglu, C. Farabet, "Convolutional Networks and Applications in Vision", Proc. International Symposium on Circuits and Systems (ISCAS'10), IEEE, (2010)

[32] A.L. Edelen *et al.*, "First Steps Toward Incorporating Image Based Diagnostics Into Particle Accelerator Control Systems Using Convolutional Neural Networks", In Proceeding of Proceedings of NAPAC2016, Chicago, IL, USA (2016)

[33] D. Vilsmeier, M. Sapinski and R. Singh, "Space-charge distortion of transverse profiles measured by electron-based ionization profile monitors and correction methods", Phys. Rev. Accel. Beams (22), 052801 (2019).

[34] T. Boltz, T. Asfour, M. Brosi, E. Bründermann, B. Härer, P. Kaiser, *et al.*, "Feedback Design for Control of the Micro-Bunching Instability based on Reinforcement Learning", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 104–107, doi:10.18429/JACoW-IPAC2019-MOPGW017

[35] A. Sanchez-Gonzalez *et al.*, "Machine learning applied to single-shot x-ray diagnostics in an XFEL", Nature Communications 8, 15461 (2017).

[36] A. Scheinker *et al.*, "Model-independent tuning for maximizing free electron laser pulse energy", Phys. Rev. Accel. Beams (22), 082802 (2019).

[37] E. Bozoki, A. Friedman, "Neural network technique for orbit correction in accelerators/storage rings", AIP Conference Proceedings 315, 103 (1994).

[38] E. Meier, G. LeBlanc, and Y. E. Tan, "Orbit Correction Studies using Neural Networks", in *Proc. 3rd Int. Particle Accelerator Conf. (IPAC'12)*, New Orleans, LA, USA, May 2012, paper WEPPP057, pp. 2837–2839.

[39] Y. Kijima, M. Mizota, K. Yoshida, K. Suzuki, "A Beam Diagnostic System for Accelerator using Neural Networks", In Proceedings of 3rd European Particle Accelerator Conference, Berlin, Germany (1992).

[40] Y. Li, R. Rainer, W. Cheng, "Bayesian approach for linear optics correction", Phys. Rev. Accel. Beams (22), 012804 (2019).

[41] T. Persson, F. Carlier, J. Coello de Portugal, A. Garcia-Tabares Valdivieso, A. Langner, E.H. Maclean, L. Malina, P. Skowronski, B. Salvant, R. Tomas, and A.C. Garcia Bonilla, "LHC optics commissioning: A journey towards 1% optics control", Phys. Rev. Accel. Beams (20), 061002 (2017).

[42] R.Tomás, M. Aiba, A. Franchi, U. Iriso, "Review of linear optics measurement and correction for charged particle accelerators", Phys. Rev. Accel. Beams (20), 054801 (2017).

[43] E. Fol, J.M. Coello de Portugal, G. Franchetti, and R. Tomás, "Optics Corrections Using Machine Learning in the LHC", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 3990–3993, doi:10.18429/JACoW-IPAC2019-THPRB077

[44] F. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825–2830, 2011.

[45] M. Aiba *et al.*, "First β-beating measurement and optics analysis for the CERN Large Hadron Collider", Phys. Rev. Accel. Beams (12), 081002 (2009).

[46] J. W. Dilly *et al.*, "An Updated Global Optics Correction Scheme", CERN-ACC-NOTE-2018-0056.

[47] W.F. Bergan, I.V. Bazarov, C.J.R. Duncan, and D. L. Rubin, "Applications of Dimension-Reduction to Various Accelerator Physics Problems", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 4060–4062, doi:10.18429/JACoW-IPAC2019-THPRB099

[48] G.E. Hinton, R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", Science, 313(5786):504 (2006).

[49] J. W. Leea, S. Chob, M. Yoon, "A neural-network method for diagnosing beam-position monitors in storage ring", Nucl. Instrum. Methods Phys. Res., Sect. A, 402 (19), pp. 14-20, 1997.

[50] E. Fol, "Detection of faulty Beam Position Monitors", presented at ICFA Beam Dynamics Mini-Workshop: Machine Learning Applications for Particle Accelerators, Menlo Park, CA, USA, 2018.

[51] E. Fol, J.M. Coello de Portugal, and R. Tomás, "Unsupervised Machine Learning for Detection of Faulty Beam Position Monitors", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 2668–2671, doi:10.18429/JACoW-IPAC2019-WEPGW081

[52] R. Calaga, R. Tomás, "Statistical analysis of RHIC beam position monitors performance", Phys. Rev. ST Accel. Beams 7, 042801 (2004).

[53] A. Langner, G. Benedetti, M. Carlà, U. Iriso, Z. Martí, J. Coello de Portugal, R. Tomás, "Utilizing the N beam position monitor method for turn-by-turn optics measurements", *Phys. Rev. Accel. Beams* vol. 19, p. 092803, 2016. doi:10.1103/PhysRevAccelBeams.19.092803

[54] A. Wegscheider, A. Langner, R. Tomás and A. Franchi, "Analytical N beam position monitor method", *Phys. Rev. Accel. Beams* vol. 20, p. 111002, 2017. doi:10.1103/PhysRevAccelBeams.20.111002

[55] M. Breunig, H. P. Kriegel, R. T. Ng, J. Sander, "LOF: identifying density-based local outliers", in *ACM sigmod record*, vol. 29 no. 2, pp. 93–104, 2002. doi:10.1145/335191.335388

[56] F. Liu, K.M. Ting, Z. Zhou, "Isolation forest", In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, 413-422 (2008).

[57] G. Azzopardi, A. Muscat, S. Redaelli, B. Salvachua, and G. Valentino, "Operational Results of LHC Collimator Alignment Using Machine Learning", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 1208–1211, doi:10.18429/JACoW-IPAC2019-TUZZPLM1