ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

APPLICATION OF MACHINE LEARNING TO BEAM DIAGNOSTICS

E.Fol[†], J. Coello de Portugal, R. Tomas, CERN, CH-1211, Geneva, Switzerland
[†]also at Johann Wolfgang Goethe-University Frankfurt, 60323 Frankfurt am Main, Germany

Abstract

Machine learning (ML) techniques are widely used in science and industry as a powerful tool for data analysis and automation. Currently, in accelerator physics ML is represented as a young research field, demonstrating mixed results in the latest attempts. The presented work is devoted to exploration of appropriate ML methods for beam diagnostics. The target is to provide an overview of ML techniques which can be applied to improve beam diagnostics and general accelerator performance. Besides the results of ML tools currently used in modern accelerators and evaluation of these tools, we also demonstrate possible concepts with the potential for further investigation and give recommendations on efficient use of ML techniques in accelerators.

HISTORICAL MOTIVATION

Traditional Optimization Techniques

Various optimization problems arise in modeling and operation of accelerators. Multi-parameter optimization can be performed using well established methods such as simplex, as it was shown at KEKB [1] applied on minimization of vertical emittance in injector linac. Another examples for the application of simplex as optimization technique for accelerators can be found in [2], where the method was applied on tuning of beam delivery system in CLIC simulations.

Linear optics corrections using optimization algorithms to find a minimum beam size with multi-parameter knobs as input were performed already in 1993 [3]. Luminosity maximization and beam lifetime are typical multivariate optimization tasks also in circular colliders [4, 5].

For light sources, such methods as random-walk optimization are being used to reduce the vertical beam size [6]. Also online optimization using various measures of accelerator performance as objective functions is being successfully applied in operation [7]. An illustrating example towards applying ML is the development of machine based optimization using genetic algorithms, which have been used as well in accelerator design [8–10].

Limitations of Traditional Methods

The described examples demonstrate successful performance of traditional optimization tools in applications on linear optics corrections and problems with limited amount of optimization targets. 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 accelerators such HE-LHC and FCC. The main limitation of traditional optimization methods is that the objective function or specific rules and thresholds have to be known. In opposite, Machine learning (ML) methods can learn from given examples without requiring explicit rules.

RELEVANT MACHINE LEARING 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 [11].

Depending on 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 known. 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.

Artificial Neural Network

Artificial Neural Networks (ANNs) are well suited for learning tasks, where data is represented by noisy, complex sensor 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 [12] 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 Adam [13] and AdaGrad [14] 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.

ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

One interesting property of the backpropagation based learning is its ability to discover useful intermediate representation inside the network. Features that are not explicitly given can be extracted using layers between input and output layers (*hidden layers*). Thus, properties of the input that are most important for the learning can be discovered. This ability is a great advantage of this method in contrast to ML techniques that use only predefined features. ANNs with many hidden layers called *deep neural networks* are able to use fewer neurons per layer and have a better generalization ability [15], however the optimization 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 [16–18]. A broad presentation of concepts and applications of ANN to particle accelerators is given in [19].

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 rules [20, 21]. Each leaf corresponds to one class representing the most appropriate class label. For the regression the leaf nodes correspond to 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 [22]. 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 correct 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 [23, 24]. 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 [25], which is based on centroid search. Another kind of clustering algorithms are the density-based algorithms such DBSCAN [26], 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.

OVERVIEW ON CURRENT APPLICATIONS

Meeting the demand of experimenters induce various challenges for accelerator design in general and in particular for beam control and diagnostics. Considering impressive results of ANNs applied in various scientific fields [27–31], among others in HEP [32] and the increase in available computational power, ML can cast the light on novel solutions for these challenges. In the following we demonstrate some ML applications currently being used in accelerator technology and ongoing research on potential ML based approaches.

Optimization and Prediction

A complex system such as an accelerator, which beam dynamics exhibits nonlinear response to machine settings can be considered as a typical ML task. ML methods are especially suited for non-linear and time-varying systems with large parameter spaces. Due to the constant increase of machine design complexity and development of new interacting systems, traditional techniques might become insufficient.

ANN based application has been successfully applied at the Linac Coherent Light Source (LCLS) to predict x-ray pulse properties using electron beam and x-ray parameters as input [33]. The method is general and can be applied at any XFEL facility to obtain fast x-ray diagnostics. ANN is able to decode complex hidden correlations between parameters obtained from slow diagnostics such as photon energy and properties measured by fast diagnostics. For the training, small samples obtained at low repetition rate are used to predict complex diagnostics at a high repetition rate.

Similar approach has been applied at TEU-FEL by training an ANN with controllable machine settings such RF power and solenoid strengths as inputs in order to produce

ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

prediction of electron beam parameters such beta functions and emittance [34].

A special kind of ANN, *convolutional neural networks* (CNN) [35] have been applied at FAST on image based diagnostic during beam operation [36]. A combination of a CNN and a feed-forward NN yields promising results for the prediction of beam parameters on simulated datasets. The model uses simulated cathode images, solenoid strengths and the gun phase as inputs and produces a prediction for various downstream beam parameters.

Extremum Seeking (ES) technique [37, 38] in combination with ML is being applied to accelerator tuning and optimization providing promising results. In this application ANN is used for fast tuning in order to obtain a close approximation for the target settings. ES is used for beam-based adaptive feedback to track the actual time-varying optimal parameter settings. The advantages of this method are the model-independent approach, the ability to perform the tuning on many coupled parameters and handle time-varying noisy data.

Another example concerning the prediction of beam parameters is the application of various ML algorithms to the problem of inferring the actual beam profile width from measured profiles that are distorted by space-charge effects [39]. The promising results obtained on simulations data show the potential of the method to be further investigated on real measurements in order to reduce cost of Ionization Profile Monitors.

Automation of Operation

Often there is a need to introduce various experiment-driven settings. In this case a large amount of free parameters has to be taken into account in order to meet the requirements and find optimal machine settings under given limitations. Only few parameters can be processed by a human at once and it is not feasible to produce forecasts taking into account all possible factors and correlations. Moreover, humans can perform differently on the tasks where the decisions can be subjective, which might lead to significantly different results. Also the automation of routine tasks could bring significant improvements into daily operation, such that the focus of operation can be transferred to complex tasks and rare events that require an expertise.

Apart from ANN, it is also possible to apply other kinds of regressors or classifiers in accelerator control such Gradient Boosting classifier [40] as it was shown in beam loss pattern classification for LHC [41]. The beam loss maps performed in controlled conditions are used in order to train a model to classify the type of losses during the LHC machine cycle.

Another example of ML based automation is the automatic alignment of collimators in SPS and LHC [42]. The method computes optimum angular settings for the collimators without human intervention. Fully automated alignment was achieved after the introduction of the ML based detection of alignment spikes in the losses which are used in order to determine if a collimator needs to be moved.

Table 1: Performance of Trained Models

Injection optics

Model	**MAE [1×10^{-5} m^{-2}]**	**Explained σ^2**
Random Forest	0.005	0.99
OMP	0.04	0.97
Neural Network	0.35	0.38

$\beta^* = 40$ cm

Model	**MAE [1×10^{-5} m^{-2}]**	**Explained σ^2**
Random Forest	0.005	0.99
OMP	0.21	0.76
Neural network	0.33	0.47

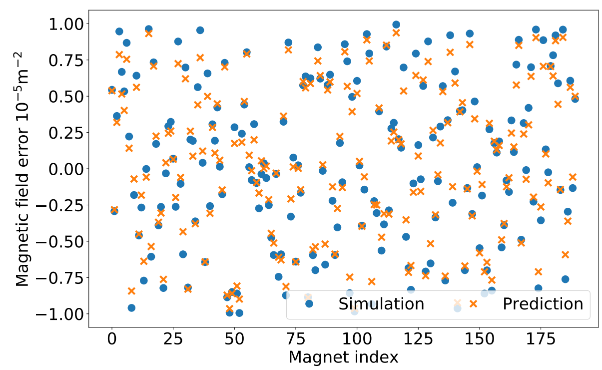

Figure 1: Random Forest prediction result on a random sample from the test set giving mean absolute error 0.02×10^{-5} m^{-2}.

Lattice Imperfections Correction

Attempts to build beam diagnostics and beam control systems using ML have been made already in the past decades [43–45]. The application described in [45] was built to detect dipole errors aiming to develop rapid commissioning. In this application, the dipole errors are obtained from the deviation of the measured beam position from the computed one. The simulations and tests have been performed on a relatively small machine (8 FODO cells, 8 BPMs), however the accuracy of the trained models decreased significantly after introduction of more than 2 dipole field defects. Anyway, given the early stage of ANN technology in that time, the obtained results have shown the potential of ML solution to be applied in beam control systems.

Recently, we applied ML on optics correction at LHC to predict the correction knobs settings required to cancel the quadrupole field errors [46, 58]. In this case, optics correction is defined as a regression problem that can be solved by training a model using past measurements or simulations and its corresponding corrections (supervised learning).

In order to create a training set, random errors are introduced into the MADX-variables that represent physical circuits. In real measurements the optics errors are caused

ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

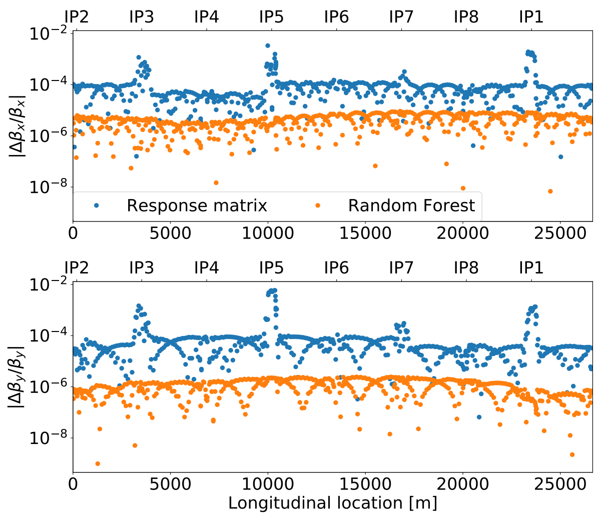

Figure 2: Expected beta-beating after applying corrections computed with linear response matrix and Random Forest regressor on simulated measurement.

by individual magnets instead of the circuits, in the following study we neglect this since the corrections can be done only using the circuit variables. The differences between the nominal model phase and the phase produced by perturbed simulations are provided as input and the corresponding errors as output of the model. Simulations data set of 10000 samples was divided into train and test set (60% and 40% respectively), each sample pair consist of 1046 inputs (number of BPMs in both planes) and 190 outputs (correction variables).

Comparison between logistic regression, ANN, Orthogonal Matching Pursuit [47] and Random Forest [48] models trained on two different optics setting are shown in Table 1. Random Forest algorithm achieves the most accurate prediction on test data set, the behavior on arbitrary sample from the test set is shown in Fig. 1.

The comparison to the traditional response matrix method as implemented in LHC [49,50] is performed on an independently simulated measurement perturbed by random errors with the absence of BPMs noise and triplet errors. The differences in obtained global optics corrections are shown in Fig. 2. Random Forest achieves an overall better correction on the given simulation and demonstrates that further enhancement in traditional method are required in order to treat the errors around interaction points (IP). The problem of the traditional method can be also related to the linear behavior of the response matrix, concerning possible nonlinearities in these regions.

Further improvements in ML based optics correction can be achieved using more realistic simulations that include other sources of optics error. The quality of the model can be significantly improved by combining different optics models in one training set in order to achieve better generalization. Potentially efficient approach for the training is *Transfer Learning*, a method where a trained model can use the learned representation to solve similar tasks after being re-trained on a very small data set [51,52]. Thus, small amount of real optics measurements and performed corrections can be used to tune the model trained on simulations to significantly increase the model quality.

Anomaly 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), cluster analysis (unsupervised learning) or applying 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 [53].

An early example on anomaly detection in beam diagnostics in storage ring (Pohang Light Source) is the application of ANN to predict the orbit at particular BPM based on measurements at other BPMs [54]. A large deviation between measured and predicted orbit should mark malfunctioning BPM.

An example for anomaly detection using cluster analysis is the detection of faulty BPMs at LHC based on harmonic analysis on turn-by-turn data [55]. The main issue regarding the problem of faulty BPMs is the appearance of unphysical data in the reconstructed optics functions. Most of the noise can be removed using traditional methods based on SVD and FFT [56] and manual cleaning, however, faulty data samples still can be observed in the optics functions. Since no labeled data is available and the data analysis has to be performed on multidimensional parameter space, clustering appears as an effective solution. The analysis as shown in Fig. 3 is performed on a three dimensional parameter space containing the betatron tune, the amplitude of the measured oscillations and the noise to amplitude ratio. The data is normalized to the range $[0, 1]$ and separated into IR and Arcs BPMs due to the different data points distribution in these regions.

Since the appearance of outliers in the data affects the computation of the mean of parameters, the algorithms based on centroids search such as K-means are not appropriate for our problem. Instead of centroid-based algorithms, DBSCAN and Isolation Forest (IF) [57] have been applied on the turn-by-turn data after harmonic analysis. Compared to the application of DBSCAN [58] better results could be achieved using IF algorithm. IF uses an ensemble of randomized decision trees averaging the path lengths, which represent the number of splits required to isolate a data point. Shorter paths are produced for anomalies. Due to the randomization and combination of several decision trees, the method performs better than single-model methods. Another advantage of IF is that the algorithm requires only the number of trees and the proportion of outliers in the data set (threshold for the decision function) as input parameter.

This method is fully integrated into optics measurements at LHC and has been successfully used during commission-

ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

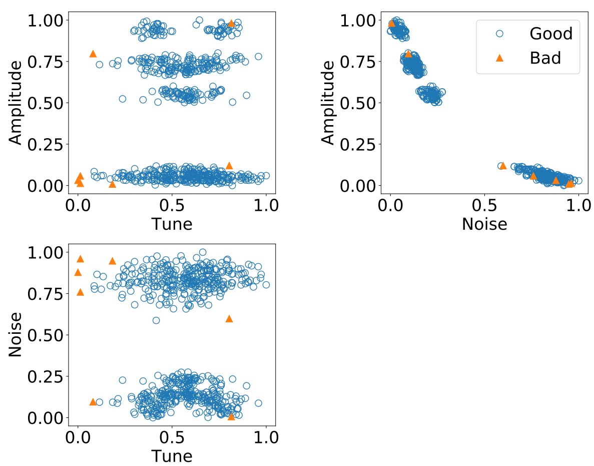

Figure 3: 2D-projection of 3D analysis with IF on arcs measurements of tune, amplitude and signal noise in horizontal plane. The data is scaled to the range [0, 1]. The triangular points correspond to BPMs marked as anomaly.

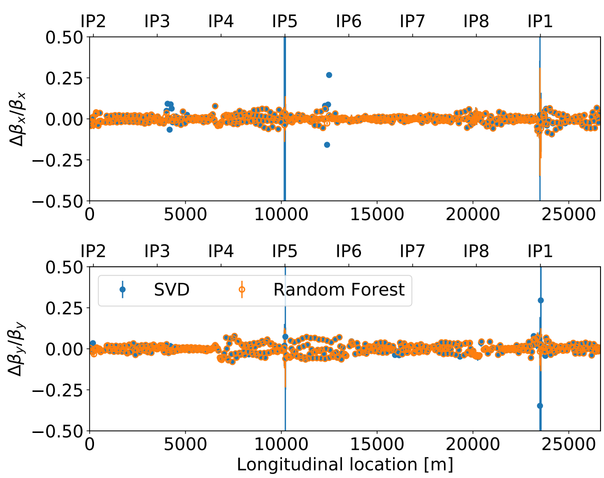

Figure 4: Beta-beating from the measurement cleaned with SVD before and after applying IF. IF decreased the number of unphysical outliers in the computed optics and significantly reduced the errorbars.

ing and machine developments under different optics settings in 2018. An illustration of results obtained during commissioning is presented in Fig. 4. The method can be further improved by introducing an adaptive decision function threshold based on corresponding normality score, that can be obtained for each sample during training process. The decision function can be then adapted depending on the numbers of BPMs lying under the threshold.

CONCLUSION

The understanding of the concepts in the field of ML and AI provides new opportunities for incorporating this discipline into accelerator technology, however the benefits in comparison to traditional tools and the efficiency of introduced ML and AI based methods need careful evaluation.

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 datasets. Most of the named 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 transfered to predict beam behavior during operation.

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 datasets 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 collimators alignment at LHC [42] is very effective and can save operational resources. Diagnostics of beam losses and beam optics corrections can also benefit from ML as it was shown in [41] and [58], since these methods rely on already existing data from dedicated measurements.

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.

Further Applications

Besides presented applications, another field where we can potentially benefit from ML are highly complex simulation tasks which demand large computational power or even are not solvable using the available resources. The predictive power of ANNs could offer an alternative solution for problems such as dynamic aperture computation avoiding costly turn-by-turn simulations. Moreover, different ML methods can be potentially used for maintenance of accelerators, such that system defects can be predicted based on deviations from expected performance and required intervention can be performed prior to actual failures [59]. Beam lifetime optimization is another task where ML techniques can be potentially applied [60]. Such methods as decision trees and autoencoders appear suitable for analysis of beam parameters correlations and their importance for beam lifetime prediction and maximization.

ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

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, USA, KEK Report No. 98-208, 1999.

[2] B. Dalena *et al.*, "Beam delivery system tuning and luminosity monitoring in the Compact Linear Collider", *Phys. Rev. ST Accel. Beams*, vol. 15, p. 051006, 2012. doi:10.1103/PhysRevSTAB.15.051006

[3] N. J. Walker, J. Irwin, and M. Woodley, "Global tuning knobs for the SLC final focus", SLAC, CA, USA, Rep. SLAC-PUB-6207, 1993.

[4] Y. Funakoshi *et al.*, "Performance of KEKB with crab cavities", in *Proceedings of the 11th European Particle Accelerator Conference*, Genoa, Italy, 2008, paper WEXG01, pp. 1893–1897.

[5] W. Fischer, J. Beebe-Wang, Y. Luo, S. Nemesure, and L. K. Rajulapati, "RHIC proton beam lifetime increase with 10- and 12-pole correctors", in *Proceedings of IPAC'10*, Kyoto, Japan, May 2010, paper THPE099, pp. 4752–4754.

[6] M. Aiba, M. Boege, N. Milas, and A. Streun, "Ultra low vertical emittance at SLS through systematic and random optimization", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 694, pp. 133–139, 2012. doi:10.1016/j.nima.2012.08.012

[7] X. Huang, J. Corbett, J. Safranek, and J. Wu, "An algorithm for online optimization of accelerators", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 726, pp. 77–83, 2013. doi:10.1016/j.nima.2013.05.046

[8] K. Tian, J. Safranek, and Y. Yan, "Machine based optimization using genetic algorithms in a storage ring", *Phys. Rev. ST Accel. Beams*, vol. 17, p. 020703, 2014. doi:10.1103/PhysRevSTAB.17.020703

[9] L. Yang, D. Robin, F. Sannibale, C. Steier, and W. Wan, "Global optimization of an accelerator lattice using multiobjective genetic algorithms", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 609, no. 1, pp. 50–57, 2009. doi:10.1016/j.nima.2009.08.027

[10] Y. Li, W. Cheng, L. H. Yu, and R. Rainer, "Genetic algorithm enhanced by machine learning in dynamic aperture optimization", *Phys. Rev. ST Accel. Beams*, vol. 21, p. 054601, 2018. doi:10.1103/PhysRevAccelBeams.21.054601

[11] T. Mitchell, *Machine Learning*, 1st ed., New York, USA: McGraw-Hill, Inc., 1997.

[12] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors", *Nature*, vol. 323, pp. 533–536, 1986. doi:10.1038/323533a0

[13] D. Kingma and J. Ba, "Adam: a method for stochastic optimization", arXiv:1412.6980, 2014.

[14] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization", *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2010.

[15] Y. Bengio, Y. LeCun, and G. Hinton, "Deep Learning", *Nature*, vol. 521, pp. 436–444, 2015. doi:10.1038/nature14539

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

[17] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, pp. 85–117, 2015. doi:10.1016/j.neunet.2014.09.003

[18] K. G. Vamvoudakis, F. L. Lewis, and Shuzhi Sam Ge, "Neural Networks in Feedback Control Systems", in *Mechanical Engineer's Handbook*, New York, USA: John Wiley, 2015. doi:10.1002/9781118985960.meh223

[19] A. L. Edelen, S. G. Biedron, B. E. Chase, D. Edstrom, S. V. Milton, and P. Stabile, "Neural Networks for Modeling and Control of Particle Accelerators", *IEEE Transaction of Nuclear Science*, vol. 63, no. 2, pp. 878–897, 2016. doi:10.1109/TNS.2016.2543203

[20] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.

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

[22] T.G. Dietterich, "Ensemble Methods in Machine Learning", in *Proceedings of the First International Workshop on Multiple Classifier Systems*, Cagliari, Italy, 2000, pp. 1–15. doi:10.1007/3-540-45014-9

[23] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning", in *Springer Series in Statistics*, New York, USA: Springer New York Inc. 2001. doi:10.1007/978-0-387-21606-5

[24] A. K. Jain, M. N. Murty, and P. J Flynn, "Data clustering: a review", *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999. doi:10.1145/331499.331504

[25] S. P. Lloyd, "Least squares quantization in PCM", *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982. doi:10.1109/TIT.1982.1056489

[26] M. Ester, H. Kriegel, J. Sander, and X. 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*, Portland, Oregon, USA, Aug. 1996, pp. 226–231.

[27] S. Chatterjee, A. Iyer, S. Avva, A. Kollara, and M. Sankarasubbu, "Convolutional Neural Networks In Classifying Cancer Through DNA Methylation", arXiv:1807.09617, 2018.

[28] M. Vidyasagar, "Machine learning methods in the computational biology of cancer", in *Proceedings of the Royal Society A*, vol. 470, no. 2167, 2014. doi:10.1098/rspa.2014.0081

[29] O. Hegazy, O. S. Soliman, and M. A. Salam, "A Machine Learning Model for Stock Market Prediction", *International Journal of Computer Science and Telecommunications*, vol. 4, no. 12, pp. 17–23, 2013.

[30] S. Hara, T. Ono, R. Okamoto, T. Washio, and S. Takeuchi, "Anomaly detection in reconstructed quantum states using a machine-learning technique", *Physical Review A*, vol. 89, no. 2, p. 022104, 2014. doi:10.1103/PhysRevA.89.022104

ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

[31] P. Graff *et al.*, "SkyNet: an efficient and robust neural network training tool for machine learning in astronomy", *Monthly Notices of the Royal Astronomical Society*, vol. 441, no. 2, pp. 1741–1759, 2014. doi:10.1093/mnras/stu642

[32] A. Aurisano *et al.*, "A convolutional neural network neutrino event classifier", *Journal of Instrumentation*, vol. 11, no. 9, p. P09001, 2016. doi:10.1088/1748-0221/11/09/P09001

[33] A. Sanchez-Gonzalez *et al.*, "Machine learning applied to single-shot x-ray diagnostics in an XFEL", *Nature Communications* vol. 8, p. 15461, 2017. doi:10.1038/ncomms15461

[34] A.L. Edelen, S. Biedron, S.V. Milton, and P.J.M. van der Slot, "Using Neural Network Control Policies for rapid switching between beam parameters in an FEL", in *Proceedings of the 38th International Free Electron Laser Conference*, Santa Fe, NM, USA, Feb. 2017, pp. 488–491. doi:10.18429/JACoW-FEL2017-WEP031

[35] Y. LeCun, K. Kavukvuoglu, and C. Farabet, "Convolutional networks and applications in vision", in *Proc. International Symposium on Circuits and Systems (ISCAS'10)*, Paris, France, Jun. 2010, pp. 253–256. doi:10.1109/ISCAS.2010.5537907

[36] A.L. Edelen *et al.*, "First steps toward incorporating image based diagnostics into particle accelerator control systems using convolutional neural networks", in *Proc. NAPAC2016*, Chicago, IL, USA, 2016, pp. 390–393. doi:10.18429/JACoW-NAPAC2016-TUPOA51

[37] A. Scheinker, "Model independent beam tuning", in *Proceedings IPAC'13*, Shanghai, China, May 2013, paper TUPWA068, pp. 1862–1864.

[38] A. Scheinker *et al.*, "Bounded extremum seeking with discontinuous dithers", *Automatica*, vol. 69, pp. 250–257, 2016. doi:10.1016/j.automatica.2016.02.023

[39] D. Vilsmeier, M. Sapinski, R. Singh, and J. W. Storey, "Reconstructing space-charge distorted IPM profiles with Machine Learning algorithms", in *Proceedings IPAC'18*, Vancouver, Canada, Jul. 2018, pp. 2099–2102. doi:10.18429/JACoW-IPAC2018-WEPAK008

[40] J.H. Friedman, "Greedy function approximation: a gradient boosting machine", *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001. doi:10.1214/aos/1013203451

[41] G. Valentino and B. Salvachua, "Machine learning applied at the LHC for beam loss pattern classification", in *Proceedings IPAC'18*, Vancouver, Canada, Jun. 2018, pp. 2020–2023. doi:10.18429/JACoW-IPAC2018-WEPAF078

[42] G. Azzopardi *et al.*, "Automatic angular alignment of the LHC collimators", in *Proceedings of ICALEPCS'17*, Barcelona, Spain, Jan. 2017, pp. 928–933. doi:10.18429/JACoW-ICALEPCS2017-TUPHA204

[43] E. Bozoki and A. Friedman, "Neural network technique for orbit correction in accelerators/storage rings", *AIP Conference Proceedings*, vol. 315, pp. 103–110, 1994. doi:10.1063/1.46759

[44] E. Meier, Y.R.E. Tan, and G. S. LeBlanc, "Orbit correction studies using neural networks", in *Proceedings of IPAC'12*, New Orleans, Louisiana, USA, May 2012, paper WEPPP057, pp. 2837–2839.

[45] Y. Kijima, M. Mizota, K. Yoshida, and K. Suzuki, "A beam diagnostic system for accelerator using neural networks", in *Proceedings of 3rd European Particle Accelerator Conference*, Berlin, Germany, Mar. 1992, pp. 1155–1157.

[46] E. Fol, "Evaluation of machine learning methods for LHC optics measurements and corrections software", Master thesis, CERN, Geneva, Switzerland, Rep. CERN-THESIS-2017-336, 2017.

[47] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit", Technion, Computer Science Department, Rep. CS-2008-08-2008, 2008.

[48] L. Breiman, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi:10.1023/A:1010933404324

[49] 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*, vol. 20, p. 061002, 2017. doi:10.1103/PhysRevAccelBeams.20.061002

[50] R.Tomás, M. Aiba, A. Franchi, and U. Iriso, "Review of linear optics measurement and correction for charged particle accelerators", *Phys. Rev. Accel. Beams*, vol. 20, p. 054801, 2017. doi:10.1103/PhysRevAccelBeams.20.054801

[51] J. J. Lim, R. Salakhutdinov, and A. Torralba, "Transfer learning by borrowing examples for multiclass object detection", in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, Granada, Spain, Dec. 2011, pp. 118–126.

[52] S. J. Pang and Q. Yang, "A survey on transfer learning", *IEEE Transactions on Knowledge and Data Engineering*, vol. 22 no. 10, pp. 1345–1359, 2010. doi:10.1109/TKDE.2009.191

[53] G.E. Hinton and R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks", *Science*, vol. 313, no. 5786, pp. 504–507, 2006. doi:10.1126/science.1127647

[54] J. W. Leea, S. Chob, and M. Yoon, "A neural-network method for diagnosing beam-position monitors in storage ring", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 402, no. 1, pp. 14–20, 1998. doi:10.1016/S0168-9002(97)01063-2

[55] 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, unpublished.

[56] R. Calaga and R. Tomás, "Statistical analysis of RHIC beam position monitors performance", *Phys. Rev. ST Accel. Beams*, vol. 7, p. 042801, 2004. doi:10.1103/PhysRevSTAB.7.042801

[57] F. Liu, K.M. Ting, and Z. Zhou, "Isolation forest", *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, Pisa, Italy, 2008, pp. 413–422. doi:10.1109/ICDM.2008.17

[58] E. Fol, F. Carlier, J. Coello de Portugal, A. Garcia-Tabares Valdivieso, and R. Tomás, "Machine learning methods for optics measurements and corrections at LHC", in *Proceedings IPAC'18*, Vancouver, Canada, May 2018, pp. 1967–1970. doi:10.18429/JACoW-IPAC2018-WEPAF062

ISBN: 978-3-95450-201-1 doi:10.18429/JACoW-IBIC2018-TUOA02

[59] M. Hebbel, private communication, Jul. 2018.

[60] L. Coyle, "Machine learning techniques applied to lifetime optimization in LHC-2016 data set analysis", https://indico.cern.ch/event/738306/.