

**68th ICFA ADVANCED BEAM DYNAMICS WORKSHOP ON HIGH-INTENSITY
AND HIGH-BRIGHTNESS HADRON BEAMS — HB2023****Optimizing dynamic aperture studies with active learning****D. Di Croce** ^{a,*} **M. Giovannozzi**^b **E. Krymova**^c **T. Pieloni**^a **S. Redaelli**^b **M. Seidel**^{a,d}
R. Tomás^b and **F.F. Van der Veken**^b^a*LPAP, École Polytechnique Fédérale de Lausanne,
BSP — Rte de la Sorge, Lausanne, Switzerland*^b*ABP, CERN,
Espl. des Particules 1, Meyrin, Switzerland*^c*SDSC, The Swiss Data Science Center,
Wasserwerkstrasse 10, Zurich, Switzerland*^d*GFA, Paul Scherrer Institut,
Forschungsstrasse 111, Villigen, Switzerland*E-mail: davide.dicroce@epfl.ch

ABSTRACT: Dynamic aperture is an important concept for the study of non-linear beam dynamics in circular accelerators. It describes the extent of the phase-space region where a particle's motion remains bounded over a given number of turns. Understanding the features of dynamic aperture is crucial for the design and operation of such accelerators, as it provides insights into nonlinear effects and the possibility of optimising beam lifetime. The standard approach to calculate the dynamic aperture requires numerical simulations of several initial conditions densely distributed in phase space for a sufficient number of turns to probe the time scale corresponding to machine operations. This process is very computationally intensive and practically outside the range of today's computers. In our study, we introduced a novel method to estimate dynamic aperture rapidly and accurately by utilising a Deep Neural Network model. This model was trained with simulated tracking data from the CERN Large Hadron Collider and takes into account variations in accelerator parameters such as betatron tune, chromaticity, and the strength of the Landau octupoles. To enhance its performance, we integrate the model into an innovative Active Learning framework. This framework not only enables retraining and updating of the computed model, but also facilitates efficient data generation through smart sampling. Since chaotic motion cannot be predicted, traditional tracking simulations are incorporated into the Active Learning framework to deal with the chaotic nature of some initial conditions. The results demonstrate that the use of the Active Learning framework allows faster scanning of the configuration parameters without compromising the accuracy of the dynamic aperture estimates.

KEYWORDS: Accelerator modelling and simulations (multi-particle dynamics, single-particle dynamics); Beam dynamics; Simulation methods and programs

*Corresponding author.

Contents

1	Introduction	1
2	Simulated samples	2
2.1	Simulating accelerator configurations	2
2.2	Simulating the DA	2
2.3	Data preprocessing	3
3	DA regressor	4
3.1	Network Architecture	4
3.2	Training and precision	4
3.3	Generalising DA prediction	5
3.4	Error estimator	6
4	Active learning framework	7
4.1	Smart sampling	8
4.2	Improving performance with synthetic data	9
4.3	Time performance	10
5	Conclusions	11

1 Introduction

The exploration of dynamic aperture (DA), which is defined as the extent of the connected phase-space region within which the dynamics of a single particle remains bounded in circular accelerators (see, e.g. ref. [1] and references therein), offers invaluable insights into the non-linear beam dynamics of non-interacting particles and the mechanisms contributing to beam losses [2]. This knowledge is crucial to optimise machine performance.

To calculate the DA numerically, one must track a multitude of initial conditions in phase space over numerous turns (at least 10^5 turns). This method imposes significant computational demands, particularly in the case of large accelerator rings such as the CERN Large Hadron Collider (LHC) [3], its luminosity upgrade HL-LHC [4], or the Future Circular Colliders under study at CERN [5, 6]. The difficulty is not in the number of initial conditions for which parallelisation can be applied, but rather in the number of turns, as no optimisation can be applied to reduce the total time required to obtain the simulation results. For these reasons, intense efforts have been devoted to the determination of analytical scaling laws that provide the evolution of the DA over time. In this way, numerical simulations can be used to successfully determine the evolution of DA well beyond the maximum simulated number of turns [7, 8].

In recent years, machine learning (ML)-based approaches have been explored to compute efficient surrogate models of DA (e.g., see [9, 10]). An alternative approach has emerged that takes advantage of deep learning (DL) to quickly and accurately predict DA for machine configurations that are unknown

to the model [11]. This was achieved by training a Deep Neural Network (DNN) on a substantial data set of simulated initial conditions. This DNN effectively captures the intricate relationship among accelerator parameters, initial conditions, and the resulting DA values. In this study, we take this integration of ML techniques a step further by incorporating an Active Learning (AL) framework. To this end, we introduce an error estimator in conjunction with the DA model to assess the uncertainty in the predictions. This addition allows the AL algorithm to conduct intelligent sampling. In other words, it can prioritise which new accelerator configurations to simulate first, guided by the magnitude of the predicted DA error. This approach aims to facilitate the rapid estimation of DA and its associated error for new accelerator parameters while simultaneously expanding the initial data set in an efficient manner, all with the added objective of enhancing the performance of the ML model.

2 Simulated samples

2.1 Simulating accelerator configurations

To train the DNN, we simulated several accelerator configurations using MAD-X [12] and the 2023 LHC lattice at injection energy (450 GeV) [13]. We varied five accelerator parameters, namely the betatron tunes Q_x, Q_y , chromaticities Q'_x, Q'_y , strength of the Landau octupoles (using the current I_{MO}) and the realisations (also called seeds) of the magnetic field errors that have been assigned to the various magnet families. Furthermore, both Beam 1 (clockwise) and Beam 2 (counter-clockwise) have been considered in these studies. We performed a random uniform grid search of the following parameters: $Q_x \in [62.1, 62.5]$ and $Q_y \in [60.1, 60.5]$, with steps of size 5×10^{-3} , $Q' \in [0, 30]$, with steps of size 2, $I_{MO} \in [-40 \text{ A}, 40 \text{ A}]$, with step size 5 A, and 60 realisations of the magnetic errors for both beams. Furthermore, we trained the model on the 2016 and 2023 LHC optics for comparison; the main difference between them is the implementation of the Achromatic Telescopic Squeezing (ATS) [14, 15] optics in the 2023 LHC configuration.

2.2 Simulating the DA

To simulate DA, we probed the phase space by tracking the initial conditions with XSuite [16, 17] for 10^5 turns. These particles are initially distributed in polar coordinates, allowing us to assess the stability limit for each angle, the so-called angular DA, i.e. the largest amplitude that survives for up to 10^5 turns. With this approach, determining the stable limit is straightforward by finding the particle with the largest amplitude that remained bounded. However, it should be noted that this method may not be as precise for concave DA regions. Furthermore, the angular DA approach offers a higher density of initial conditions for smaller radii, which is particularly valuable when calculating loss rates and considering a Gaussian particle distribution. Unlike Cartesian coordinates, where the coordinates are evenly distributed in both x and y, polar coordinates provide a higher particle density at smaller radii. For this reason, probing with polar coordinates improves the precision of loss-rate measurements.

To speed up tracking, we performed an initial scan of a set of conditions uniformly distributed in 8 polar angles in $[0, \pi/2]$ and 33 radial amplitudes in $[0.0, 20\sigma]$, to identify the value of the angular DA for a given angle, as well as the first amplitude where the particle does not survive more than 10^3 turns (limit of the fast-loss zone). Then a finer scan of the initial conditions was performed in an amplitude range between the stable zone -2σ and the fast loss zone $+2\sigma$, with radial steps of 0.06σ along 44 polar angles in $[0, \pi/2]$.

This procedure enables for a more detailed scan in phase space, specifically focusing on the region of interest. This approach not only enhances the precision of the angular DA but also accelerates tracking by an average factor of three compared to probing the whole phase space, as established empirically. An example of the results of these computations in the x - y space is shown in figure 1 for a specific accelerator configuration. It is important to note that the accelerator configuration depicted in the graph does not utilise the operational tunes, which would have provided a DA of about 7σ .

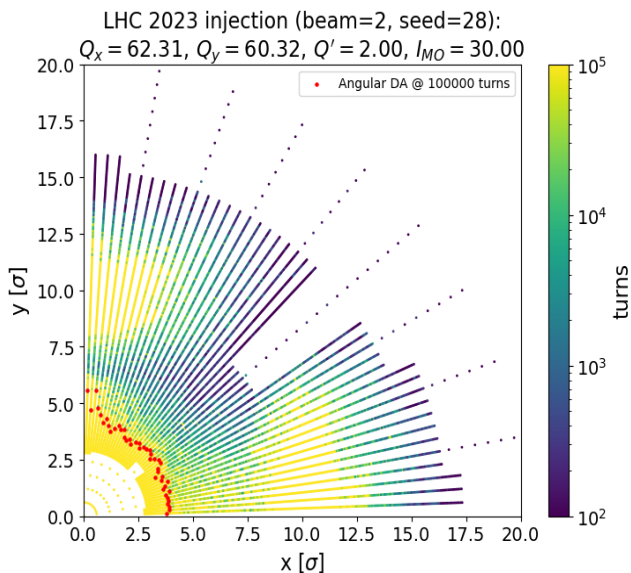


Figure 1. Stability time for a distribution of initial conditions used to compute the DA for a specific accelerator configuration. The angular DA at 10^5 turns is shown in red.

In addition, our objective includes understanding the evolution of beam losses across various accelerator configurations. Our approach involves studying how the DA changes while monitoring the survival of the particles at different numbers of turns. For this reason, the stability times have been binned using 12 distinct stability-time limits, namely 1, 5, 10, 20, 40, 50, 60, 70, 80, and 100 thousand turns.

The number of samples used for our model is calculated by multiplying three components: the number of accelerator configurations present in the data set, the number of angles used to probe the phase space (44), and the number of bins (12) used to probe the evolution of the stability time, which gives a substantial 528-fold expansion in the data set size, significantly enhancing the likelihood of achieving convergence during model training, while also allowing us to monitor the shape of the stability limits and its evolution. From this sample size, 10% of the samples were used for validation and 10% to test the performance of the model.

2.3 Data preprocessing

To gain more insight into beam dynamics, we added other variables in addition to those used to differentiate the accelerator configuration in the simulations, which are calculated by MAD-X and PTC [18]. We considered seven anharmonicity coefficients, i.e. the amplitude detuning terms up to the second order, to take into account the impact of non-linear effects on the beam dynamics. Furthermore, we included the maximum value of the Twiss parameters α and β , and the phase advance ($\mu_{x,y}$) between IP1 and IP5 to take into account the impact on the optical parameters from the nonlinear field errors.

Moreover, we carried out standardisation on the continuous input variables, excluding beam and seed variables. This involved normalising the distributions of these variables using their respective averages and standard deviations. This crucial step guarantees uniform scaling across all input features, facilitating faster model convergence and model stability. As a result, this leads to improved model performance and interpretability. Furthermore, we restrict the angular DA values to a maximum of 18σ . Values exceeding this threshold are considered outliers in a distribution ranging from 0 to 20σ . This restriction is implemented to prevent the influence of extremely high angular DA values on the regressor, as discussed in [19].

3 DA regressor

3.1 Network Architecture

Similarly to our previous study [11], we used a Multilayer Perceptron (MLP) [20] as the architecture for our surrogate model. This MLP is responsible for examining accelerator parameters, the polar angle, and tracked turns to regress the angular DA. It is essential to acknowledge that MLPs have some limitations; one of them is their lack of enforceability of constraints on particle survival time, which means that they cannot inherently ensure that angular DA decreases as we extend particle tracking over more turns for the same machine configuration. On the contrary, recurring neural network (RNN) architectures have the ability to address this issue, but require careful convergence [21], making MLPs an attractive choice when we prioritise faster and more straightforward model training. Another disadvantage of MLPs is their lack of geometry constraints, which means that the prediction of the angular DA is independent of neighbouring values. In contrast, image-based networks can use neighbouring information [22]. However, they require additional simulated sets of accelerator configurations to achieve convergence.

The network was developed using the TensorFlow library [23]. Architecture and hyperparameters were optimised by random search with the Keras Tuner framework [24]. The best model consists of four hidden layers with 2048, 1024, 512, and 128 nodes, respectively, and employs the ReLU activation function [25]. The dropout of 1% was implemented between the hidden layers together with the L2 regulariser in the hidden layers to improve the validation performance and avoid overfitting. Moreover, we included a concatenation layer after the MLP to explicitly incorporate the discrete variables (beam and seeds) into the model’s architecture and to avoid interpolation on them. The structure of our network is shown in figure 2 (left).

3.2 Training and precision

In our initial training, we trained with 2023 data corresponding to about 15 000 different accelerator configurations, totalling approximately 8 million samples. The loss used for the regressor is the Mean Absolute Error (MAE) function that is trained with the NADAM optimiser [26]. The initial learning rate is 5×10^{-5} and is halved every 10 sequential epochs if the validation loss is not improved. We found that training for 364 epochs is sufficient to validate the loss convergence.

The MAE of the angular DA regressor is 0.34(0.34) beam σ for the test (train) data set, and the mean absolute percentage error (MAPE) is 11.91(11.50), respectively. Furthermore, the analysis of the angular DA with the 2D histogram in figure 2 (right) reveals that the model performs well for most of the data points, with a tight cluster around the diagonal line, indicating accurate predictions.

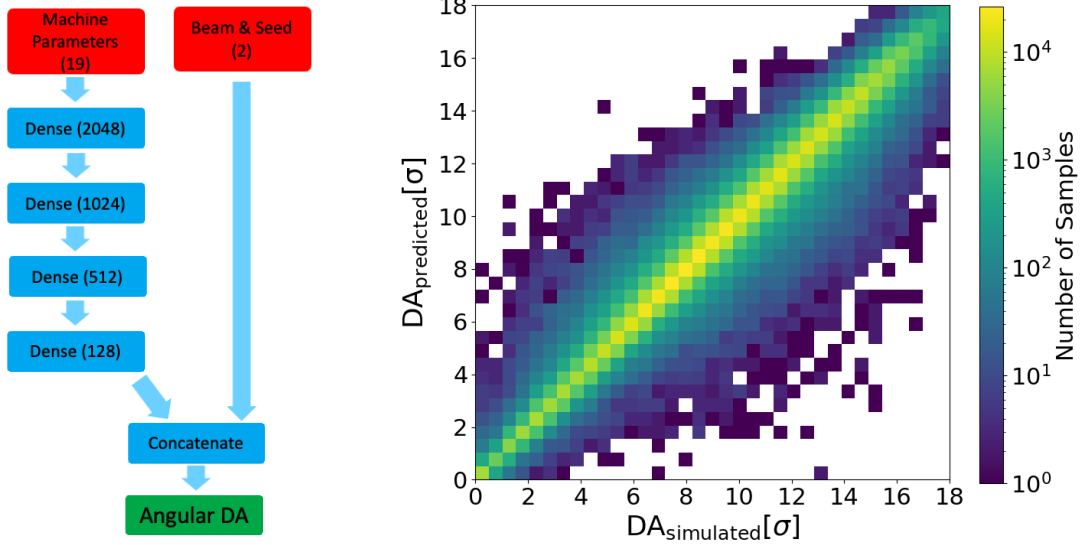


Figure 2. Left: diagram of our network architecture. Right: predicted angular DA as a function of the expected angular DA values for the test data set.

3.3 Generalising DA prediction

We want to ensure that the model can adapt to different optical configurations, which is a key requirement, especially in scenarios where the optics are unknown, such as in the design phase of a new accelerator. For this reason, we select physics variables independent of the machine optics to learn the pertinent beam dynamics features, to reduce the need for extensive retraining, and also to provide rapid adaptability to new optical configurations. To test for this capability, we trained with data from 2023 LHC optics and tested on 2016 (non-ATS) LHC optics. Both test samples have the same size (2000 configurations each).

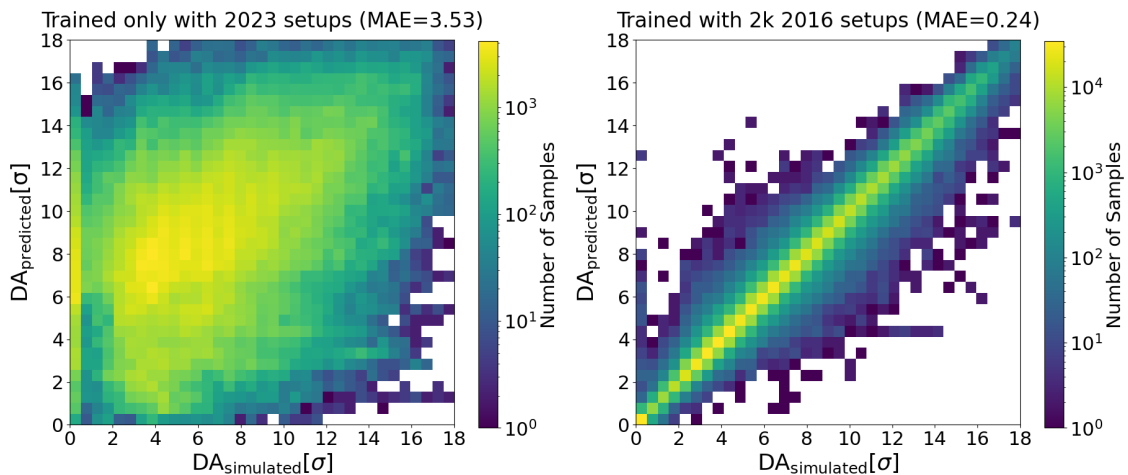


Figure 3. Predicted angular DA as a function of the expected angular DA values of the 2016 test data set for the model trained only with 2023 data (left) and that trained also with 2000 configurations of the 2016 optics (right).

It turned out that the model trained with the 2023 setup is not predictive of the 2016 test data set, as indicated by its MAE = 3.53 beam σ and the weak correlation between predicted and simulated

values shown in the 2D histogram in figure 3 (left). The main reason for this is that several input variables have a different range of values for the two data sets, such as the anharmonicities and the phase advance. This causes the model to perform an inaccurate extrapolation for these variables in the 2016 data set. Unfortunately, DNNs perform well for interpolations, but, due to their non-linear nature, not for extrapolations. An effective approach to enhance their performance in such scenarios involves the integration of adaptive feedback mechanisms, as demonstrated in recent studies (e.g., see [27, 28]). Another strategy to mitigate the extrapolation issue involves enhancing the training dataset with additional data to encompass the entire range of input variable distributions. With this intention, we have incorporated the 2016 configurations into our training dataset.

Table 1 indicates the key performance figures of the model when trained with only 2023 configurations, and with the addition of 1000, 1500 or 2000 configurations from the 2016 optics, for the 2023 (500 setups) test data set, for the 2016 (500 setups) test data set, and for both test data sets together (overall). We found out that with 2000 configurations from 2016 optics, we improved the MAE in the 2023 test to $0.33 \text{ beam } \sigma$, but more importantly, the model became predictive in the 2016 test data set with an MAE of $0.24 \text{ beam } \sigma$, as shown in figure 3 (right).

Table 1. Test MAE of the model trained with different amount of 2016 configurations.

Data Set	Only 2023	+1k 2016 setups	+1.5k 2016 setups	+2k 2016 setups
2023 test	0.34σ	0.37σ	0.35σ	0.33σ
2016 test	3.53σ	1.64σ	0.76σ	0.24σ
Overall	1.94σ	1.00σ	0.54σ	0.29σ

This ability of the model to require only a small fraction of the 2016 data to achieve predictiveness in both optics is attributed to its capacity to extract universal physics features from accelerator variables that govern the beam dynamics. This capability of rapidly adapting to new optical configurations holds immense promise for the design and optimisation of future accelerators, underscoring the critical role of surrogate models.

3.4 Error estimator

We can incorporate dropout during inference, which will randomly set some of the hidden units of the network during forward propagation to zero, to exhibit variability in the angular DA prediction. The variability in predictions introduced by the dropout can subsequently serve as a basis for estimating uncertainty; this technique is known as the Monte Carlo (MC) dropout [29].

Specifically, we introduced a 1% dropout rate between the first two hidden layers during inference and conducted 128 inferences for each angular DA prediction. We then consider two standard deviations (std) of these predictions as the angular DA uncertainty. The dropout rate, the number of std and the number of inferences chosen to predict the error were tuned to minimize the discrepancy between the true error (actual error in the prediction of the angular DA) and the one predicted with the MC dropout. The MAPE distributions of the true and predicted errors are shown in figure 4.

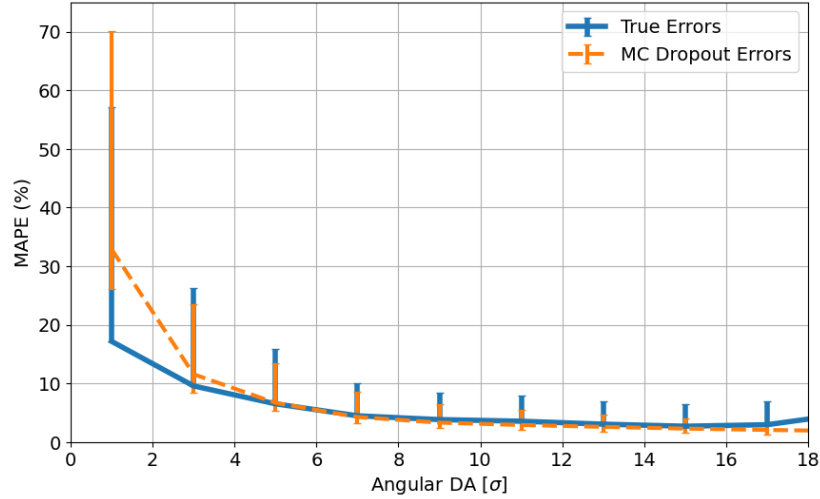


Figure 4. MAPE of the true (blue) and MC dropout (orange) errors of 2023 test data set as function of the simulated angular DA.

4 Active learning framework

Through the integration of the angular DA regressor and the error estimator, the AL framework becomes able to make informed sampling decisions based on the magnitude of the associated error. The pipeline of the AL framework is illustrated in figure 5.

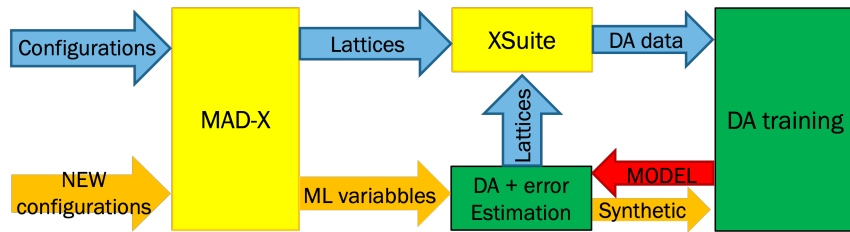


Figure 5. Pipeline of the AL framework. In yellow are the full simulation/tracking algorithms and in green are the ML algorithms. The blue arrows are the full simulated data, the orange arrows are the AL data, while the red arrow indicates the angular DA model.

In particular, when the predicted errors are small, the framework can confidently rely on the predictions of the DA regressor for a study or even to generate a synthetic data set. This preference is due to the impressive precision rate of 91.8% and a negligible fallout rate of 0.1% of our error estimator when selecting 824 configurations with Absolute Percentage Error (APE) below 10% (which is the error divided by the angular DA) from 10 000 accelerator configurations. Examples of angular DA and error predictions carried out by our AL framework with APE below 10% for four accelerator configurations (available only in the test data set) are shown in figure 6.

On the contrary, when the predicted error is significant for an accelerator configuration, the system will opt to execute the entire simulation and tracking process to generate new data, allowing the ML model to learn the features of this new accelerator configuration. Interestingly, the error estimator has an impressive accuracy rate of 99.2% and a precision rate of 98.7% when selecting 113 configurations

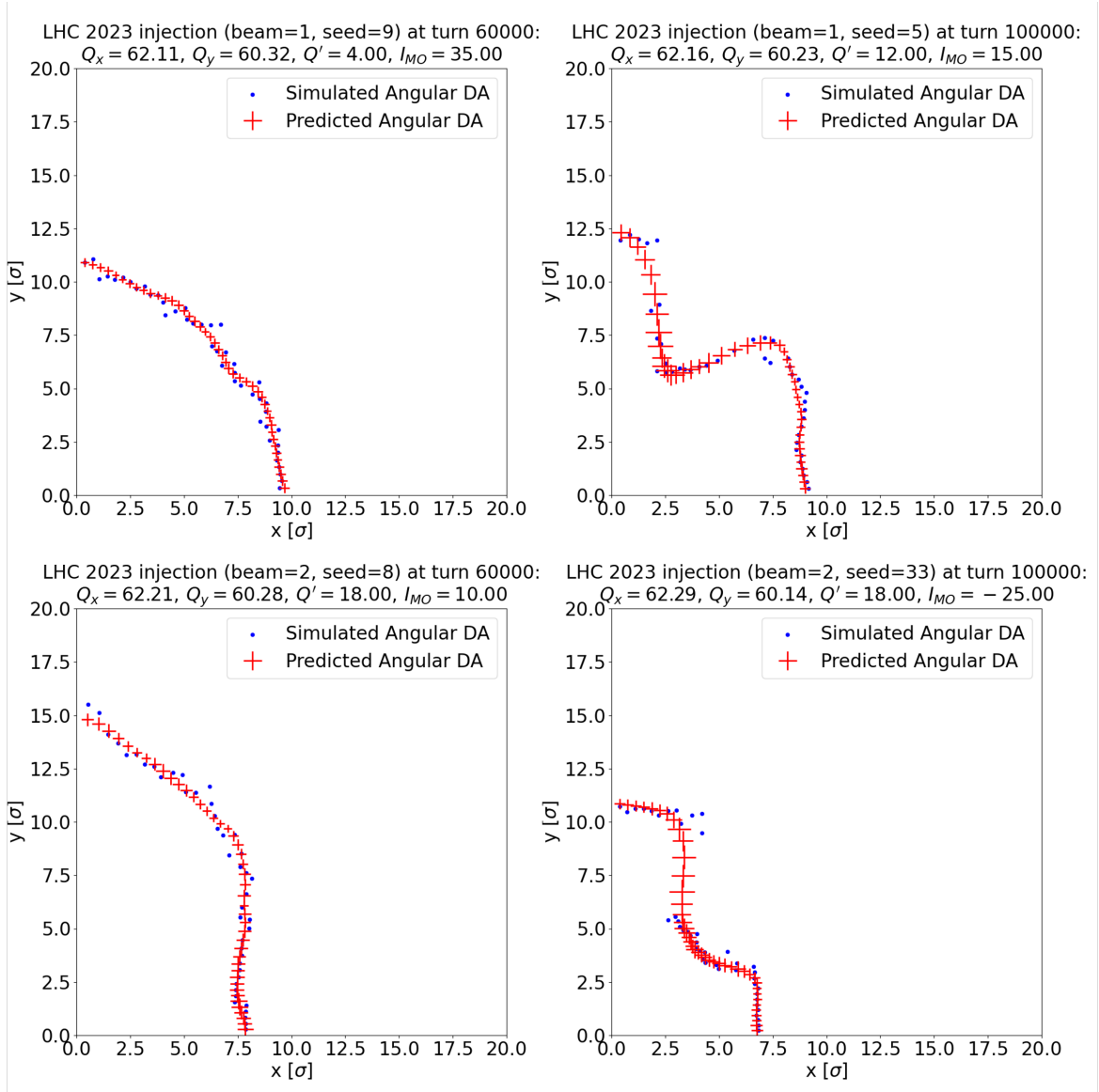


Figure 6. Simulated and predicted angular DA four configurations of the test data set (predicted APE lower than 10%).

with APE that exceed 200% of the 10 000 configurations. By giving priority to configurations with higher predicted errors, the framework efficiently determines the order in which to simulate various accelerator configurations, a strategy referred to as smart sampling.

4.1 Smart sampling

To assess the efficacy of smart sampling, we compared the performance of the model when trained with two different data sets, both derived from the 2016 LHC optics. The first data set was generated through a random search of machine configuration input, while the second was created using our innovative active learning framework. For this evaluation, we assembled a data set consisting of 5500 machine configurations, all obtained through random searches.

Among these, 4000 configurations (referred to as the initial training data set) were used to train the initial DA regressor, which in turn played a pivotal role in the configuration of the error estimator within the active learning framework. Another 1000 configurations, combined with the initial 4000, were employed to assess the performance of the model trained on random search inputs (referred to as the “Random” train data set). The remaining 500 configurations (referred to as the random testing data set) were reserved to assess the performance of the model.

To execute the smart sampling approach, our AL framework meticulously analysed a set of 100 000 configurations. From this extensive analysis, 1100 configurations with the highest predicted errors were selected, which were subsequently subjected to full simulation using MADX and XSuite. Among these selected configurations, 1000 were combined with the initial data set, collectively forming what we refer to as the smart training data set with 5000 configurations. The remaining 100 configurations were reserved as a distinct test data set, labelled the smart test data set. The performance results of the models trained with the initial, random and smart data sets are presented in table 2.

Table 2. Test MAE of the model trained with different data sets.

Data Set	Initial train	Random train	Smart train
Random test	0.22σ	0.21σ	0.21σ
Smart test	1.73σ	1.74σ	1.70σ

It is worth noting that the performance of the model on the smart test data set is lower than on the random test data set. This difference arises due to the inherent complexity of the smart test data set, which translates into higher predicted errors. In addition, the model trained with the smart training data set exhibits similar performance in the random test data set but shows an improvement in MAE when tested in the smart test data set, indicating an improvement in setups with challenging predictions. These findings indicate the effectiveness of the smart sampling approach, which allows us to explore machine configurations more efficiently where the model has not yet learnt the features of the underlying physics. This capability may potentially shed light on the chaotic behaviour of certain parameter combinations.

4.2 Improving performance with synthetic data

We evaluated the effectiveness of enhancing the precision of the DA Regressor with a synthetic data set generated by our AL framework. For this purpose, the framework predicted the angular DA and its associated error for 10 000 accelerator configurations of the 2016 LHC optics, and selected 1000 configurations with the lowest predicted APE, ensuring values lower than 2%. Variables associated with these chosen configurations, in conjunction with the predicted angular DA values, were used to construct the synthetic data set. This synthetic data set was subsequently merged with our existing 2016 simulated data set with 5000 configurations. The combined data set was then used to retrain our model. As a result, the MAE of the test data set exhibited a slight improvement, decreasing from 0.21 beam σ for the model trained with initial 5000 configurations to 0.20 beam σ for the model trained with synthetic data, compromising 6000 setups.

It is important to note the advantages of this approach in our workflow. One significant benefit is the ability to automate learning, which allows our model to refine its prediction while waiting for the acquisition of new simulated data. However, this improvement will inevitably reach a saturation point

after a few iterations if no new simulated data is introduced. As such, despite this method offering valuable ongoing refinement, it is essential to maintain a balance between automatic learning and the incorporation of fresh data to maintain the enhancement of model performance.

In future studies, in addition to considering more simulation runs by creating synthetic data, we would like to take our existing simulation data and add some normally distributed, zero-mean random noise to the input data. This approach aims to enhance the robustness of the model without altering the angular DA values, ensuring that the information remains accurate. By introducing this noise, we can provide our model with a wider range of input variations to improve its adaptability to real-world scenarios.

4.3 Time performance

The timing performance to predict the angular DA and its error for a batch of 1024 samples is 140 s (1.41 ms per inference) using a Titan V GPU [30], utilising 48 AMD Ryzen Threadripper 2970WX CPU cores for data loading. Taking into account a single accelerator configuration (12 bins for the survival time \times 44 angles), this results in an inference of 0.75 s/accelerator configuration.

The improvement in computational efficiency of our approach is a notable achievement. XSuite tracking, combined with the HTCondor system [31] which can handle around 1000 jobs concurrently, takes a waiting time about 30 h to track particles from 1000 different configurations; equivalent to 107 s/accelerator configuration. This figure slightly increases to 147 s/accelerator configuration for 10 000 configurations (roughly 17 days).¹ Therefore, our model predictions, once trained, can be up to 200 times faster than traditional tracking methods; and with newer GPUs such as the A100 [32], more than 2000 times faster.

It is important to note that both the AL framework and Xsuite rely on input from MAD-X, a process that takes about 6.3 s/accelerator configuration when executed on the HTCondor system (equivalent to approximately 17 h to run 1000 MAD-X simulations in parallel). When accounting for this simulation time, which remains consistent for both the AL framework and Xsuite, the speed-up factor for the AL framework decreases to 22 times faster than traditional simulations (MAD-X+Xsuite).

Table 3. Time, performance and speed-up factor to generate different size of data sets with MAD-X/Xsuite and AL framework.

Data set size	2500	5000	10 000	15 000
DA Regressor MAE	0.45σ	0.42σ	0.38σ	0.34σ
Time with MAD-X + Xsuite	109 h	212 h	425 h	612 h
Time with AL framework	-	117 h	126 h	136 h
Speed-up factor	-	1.8	3.4	4.5

Finally, it is important to take into account the time required for model training, which totals approximately 3 h over 350 epochs. Furthermore, we must acknowledge the need for an initial data set for training purposes. Our findings indicate that achieving convergence requires a minimum of 2500 configurations (MAE of 0.45 for beam σ), a process that consumes approximately 106 h. In light of this, the speed-up factor becomes more pronounced as the ratio of predicted to simulated data

¹This increase was the result of a decrease in user priority rank due to continuous use of the HTCondor system.

increases. However, the improvement in angular DA precision is most noticeable when the initial simulated data set is substantial, as illustrated in table 3.

5 Conclusions

In this paper, we present an innovative ML tool to study the beam dynamics in circular accelerators. Using DL networks, it was possible to accurately model and predict the DA of the CERN LHC with different optical configurations. One of the key takeaways of the DA regressor is the ability to extract physics-relevant features, enabling us to predict DA solely from accelerator variables. This approach not only reduces the need for extensive retraining but also provides rapid adaptability to new optical configurations, a crucial advantage for designing new accelerators where optics may be unknown. In addition, the DA model is a powerful tool for the preservation of knowledge. By capturing and storing knowledge of physics, our models eliminate the redundancy of repetitive simulations, serving as a valuable resource for machine optimisation.

The effectiveness of knowledge preservation is significantly enhanced through the AL framework, which plays a crucial role in accelerating the learning process. In cases where accurately predicting the DA was not possible, particularly in challenging accelerator configurations, the smart sampling strategy enabled us to first simulate these configurations. By utilising tracking tools to assess the chaotic nature of certain initial conditions, this approach significantly reduced the time required to gain insight into such features. In addition, the AL framework can generate synthetic data sets to further refine the precision of our DA model while waiting for new simulated data. Lastly, our framework has achieved outstanding results in conducting ultrafast simulations. This remarkable speed-up proves particularly advantageous for optimising existing machines and designing future accelerators.

Acknowledgments

This work is partially funded by the SDSC project C20-10 and with the support of the Swiss Accelerator Research and Technology programme [33].

This article is an updated version of the HB 2023 conference proceeding, to be published in JACOW under CC BY 3.0 license.

References

- [1] E. Todesco and M. Giovannozzi, *Dynamic aperture estimates and phase space distortions in nonlinear betatronic motion*, *Phys. Rev. E* **53** (1996) 4067.
- [2] M. Giovannozzi, *Proposed scaling law for intensity evolution in hadron storage rings based on dynamic aperture variation with time*, *Phys. Rev. ST Accel. Beams* **15** (2012) 024001.
- [3] O.S. Bruning et al., *LHC Design Report Vol.1: The LHC Main Ring*, CERN-2004-003-V1, CERN, Geneva (2004) [[DOI:10.5170/CERN-2004-003-V-1](https://doi.org/10.5170/CERN-2004-003-V-1)].
- [4] G. Apollinari et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN-2017-007-M CERN, Geneva (2017) [[DOI:10.23731/CYRM-2017-004](https://doi.org/10.23731/CYRM-2017-004)].
- [5] FCC collaboration, *FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2*, *Eur. Phys. J. ST* **228** (2019) 261.

- [6] FCC collaboration, *FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3*, *Eur. Phys. J. ST* **228** (2019) 755.
- [7] M. Giovannozzi, W. Scandale and E. Todesco, *Dynamic aperture extrapolation in presence of tune modulation*, *Phys. Rev. E* **57** (1998) 3432.
- [8] A. Bazzani et al., *Advances on the modeling of the time evolution of dynamic aperture of hadron circular accelerators*, *Phys. Rev. Accel. Beams* **22** (2019) 104003 [[arXiv:1909.09516](https://arxiv.org/abs/1909.09516)].
- [9] M. Schenk et al., *Modeling Particle Stability Plots for Accelerator Optimization Using Adaptive Sampling*, *JACoW I PAC* **2021** (2021) 1923.
- [10] M. Casanova, B. Dalena, L. Bonaventura and M. Giovannozzi, *Ensemble reservoir computing for dynamical systems: prediction of phase-space stable region for hadron storage rings*, *Eur. Phys. J. Plus* **138** (2023) 559 [[arXiv:2301.06786](https://arxiv.org/abs/2301.06786)].
- [11] D. Di Croce et al., *Accelerating Dynamic Aperture Evaluation Using Deep Neural Networks*, *JACoW I PAC2023* (2023) WEPA097.
- [12] *MAD — Methodical Accelerator Design*, <https://mad.web.cern.ch/mad/>.
- [13] R. Tomás et al., *Optics for Landau damping with minimized octupolar resonances in the LHC*, in the proceedings of the 68th ICFA ABDW on High-Intensity and High-Brightness Hadron Beams, Geneva, Switzerland, 9–13 October 2023.
- [14] S. Fartoukh, *Achromatic telescopic squeezing scheme and application to the LHC and its luminosity upgrade*, *Phys. Rev. ST Accel. Beams* **16** (2013) 111002.
- [15] S. Fartoukh et al., *Achromatic telescopic squeezing scheme and by-products: From concept to validation*, *Phys. Rev. Accel. Beams* **24** (2021) 021002.
- [16] *Xsuite documentation*, <http://xsuite.web.cern.ch>.
- [17] G. Iadarola et al., *Xsuite: an integrated beam physics simulation framework*, in the proceedings of the 68th ICFA ABDW on High-Intensity and High-Brightness Hadron Beams, Geneva, Switzerland, 9–13 October 2023 [[arXiv:2310.00317](https://arxiv.org/abs/2310.00317)].
- [18] *MAD-X-PTC documentation*, http://mad.web.cern.ch/mad/madX/doc/usrguide/ptc_general/ptc_general.html.
- [19] R.D. Cook, *Influential Observations in Linear Regression*, *J. Am. Stat. Assoc.* **74** (1979) 169.
- [20] Y. LeCun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (2015) 436.
- [21] Y. Bengio, P. Simard and P. Frasconi, *Learning long-term dependencies with gradient descent is difficult*, *IEEE Trans. Neural Networks* **5** (1994) 157.
- [22] A. Krizhevsky, I. Sutskever and G.E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou and K. Weinberger, eds., vol. 25, Curran Associates, Inc., 2012.
- [23] M. Abadi et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, [arXiv:1603.04467](https://arxiv.org/abs/1603.04467).
- [24] Keras Team, *Keras Tuner*, <https://github.com/keras-team/keras-tuner>.
- [25] V. Nair and G.E. Hinton, *Rectified linear units improve restricted boltzmann machines*, in *Proceedings of ICML*, vol. 27 (2010), pp. 807–814.
- [26] T. Dozat, *Incorporating Nesterov momentum into Adam*, in the proceedings of the 4th International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.

- [27] A. Scheinker et al., *Demonstration of Model-Independent Control of the Longitudinal Phase Space of Electron Beams in the Linac-Coherent Light Source with Femtosecond Resolution*, *Phys. Rev. Lett.* **121** (2018) 044801.
- [28] A. Scheinker, F. Cropp and D. Filippetto, *Adaptive autoencoder latent space tuning for more robust machine learning beyond the training set for six-dimensional phase space diagnostics of a time-varying ultrafast electron-diffraction compact accelerator*, *Phys. Rev. E* **107** (2023) 045302.
- [29] Y. Gal and Z. Ghahramani, *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*, in the proceedings of the 33rd International Conference on International Conference on Machine Learning, New York, New York, U.S.A., 20–22 June 2016, pp. 1050–1059 [[arXiv:1506.02142](https://arxiv.org/abs/1506.02142)].
- [30] NVIDIA Corporation, *NVIDIA TITAN V* (2017).
- [31] CERN HT-Condor documentation, <https://batchdocs.web.cern.ch/local/submit.html>.
- [32] NVIDIA Corporation, *NVIDIA A100* (2020).
- [33] Accelerator Research and Technology programme — CHART, <http://www.chart.ch>.