



Overcoming the challenges posed by beam dynamics measurements at the CERN LHC using ML techniques

Syed Owais Bukhari, S. S. Ashraf, Jawed Ahmad

School of Engineering Science and Technology, Jamia Hamdard, New Delhi-62, India
owaisbukhari123@gmail.com, shahabash@gmail.com

Abstract

As the field of high-energy particle physics advances into newer realms, the application of machine learning and deep learning techniques becomes even more important for this field than ever before. Handling higher dimensional data and most complex problems is difficult without the application of machine learning. The application of machine learning (ML) has enabled us to acquire deep insights into the accelerator physics and beam dynamics studies at the Large Hadron Collider (LHC). This has literally opened a new door of applications arising from beam measurements and particle tracking simulations. This paper is an attempt to examine the commissioning of beams in the collimation system and introduce a methodology to reduce beam losses. The research proposed here also seeks to detect collective beam instabilities and heating spikes from pressure reading samples. The paper draws a roadmap to not only revisit LHC particle physics from the lens of machine learning but also serves as a gateway to future developments in this field.

Keywords: LHC, Machine Learning, Supervised learning, Beam dynamics, Collimation system.

1. Introduction

The process of constructing a mathematical model on the basis of training data that can be used to determine the cause of an event is the prime goal of machine learning. Machine learning happens to be a sub-branch of artificial intelligence and incorporates under its umbrella different types of algorithms like supervised, unsupervised, and reinforcement learning [1, 2]. The techniques used in machine learning range from classification to clustering and regression to dimensionality reduction [3]. Training a mathematical model using machine learning techniques involves a hierarchy of steps. The first step is the collection of data and its curation [4]. This is followed by feature engineering and feature selection which reduces the dimensions of the problem in consideration. After this, model hyperparameter optimization is done, and finally

[5,6] the performance of the model during run time is evaluated.

Labeled data is used to train machine learning algorithms in such a manner that the output is as close to their predicted value as possible. The use of reinforcement learning is generally preferred while working with sequential decision-making tasks [7,8].

Significance of this work

The explosion of big data has led to the development of complex machine learning techniques like deep learning that uses convolutional neural networks to perform feature engineering and extraction. Of late, machine learning has carved itself into the domain of particle physics comprehensively. The earliest attempts to use machine learning techniques for particle identification in the Large Hadron Collider have been made years ago [9,10]. However, only a few of these attempts have met with substantial success. Talking about accelerator physics, the

need for machine learning and deep learning techniques is becoming a scientific necessity. This is what is culminating in the holistic study of particle physics of the Large Hadron Collider with the help of machine learning and deep learning techniques [11,12]. The complex operations and the dynamic configurations that are associated with the Large Hadron Collider generate large sets of data which are then collected and analyzed for particle identification, beam dynamics configuration, and optical configuration [13,14]. This suggests that machine learning can open new vistas for research in the domain of the Large Hadron Collider.

The research presented in this paper is also crucial for the development of a Hadron Collider in the future [15,16].

2. Measurements about LHC

A brief sketch of LHC ring components is described in Fig 1. In the first phase, protons are accelerated along with the ions with an energy of about 450 electron volts [17]. This process that is described in the figure shows segmented symmetry. It can be observed from the figure that the instrumentation device that produces the iron beam is located parallel to the RC system [18]. In the past, many efforts have been put into the design of linear optics and the subsequent measurements have led to the boosting of the experiments that were performed at the Large Hadron Collider [19]. However, some issues are still present which need to be resolved. For instance, the recognition of beam position monitor reading is still a challenge and research in this direction is in its initial stage. This is one area where machine learning has a point to prove [20, 21].

The correlation between machine learning and optics measurements can be explored when it comes to tracking beam trajectory and particle identification in the Large Hadron Collider. Supervised learning and unsupervised learning are prime methods used in this research. With the help of supervised learning techniques, the regression model has been constructed to enable the identification of individual magnet errors from related optics perturbations [22,23]. The techniques used today (other than machine learning) can determine the strength settings of circuits which means the deviation of optics can be computed from the design configuration. The initial set of results obtained

in the present study shows that regression models are pivotal for predicting quadrupole errors [24,25].

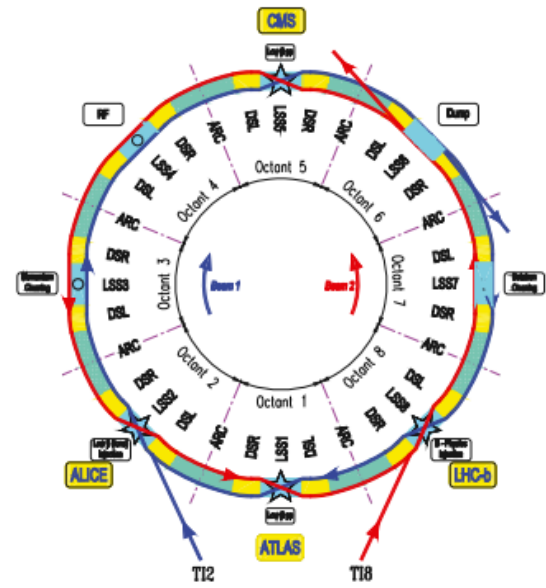


Fig. 1- A depiction of various rings of the Large Hadron Collider

It is found that supervised learning may also be applied for optics measurement which allows us to detect faulty signals. It is important to do away with BPM signals as these are the main source of outliers in the final observation [26,27]. One more important observation that is to be mentioned here is that the application of anomaly detection techniques to LHC paves the way for the analysis of informative data sets [28,29].

3. Commissioning of beams in the collimation system

The particle detection system at the Large Hadron Collider is not immune to beam losses. This can have drastic consequences on the overall studies and make the final results redundant [30,31]. Hence, there arises an immediate need to protect the collimation system from beam losses. To achieve this, an appendage of about a hundred kilometers has been designed with a radius of 27 kilometers and this commissioning system is designed in a parallel block [32]. A degree of freedom is also allowed as the corners of this system can be moved by about 400 degrees. Accuracy of six micrometers has been achieved around the circulating beam and the operational gap, in this case, has been 1 millimeter [33]. The Large Hadron

Collider collimation system is dependent upon multistage transverse settings. These settings are dependent upon the alignment of the beam-based system. The extension of the collimator system and the related components are used for measuring the loss of the beam signal [34,35]. A sophisticated beam loss monitoring device enables us to detect these beam losses. The perfect alignment of the collimator system is achieved when both jaws are aligned around the beam [36]. At this point, the graph shows a spike pattern as its output shown in Fig 2. For determining the size of the beam, the calorimeter provisions are compared with a reference beam signal obtained by similar measurements [37,38]. This is where machine learning techniques can help in deriving insights from these beam signal outputs. Ranging from particle identification to the reconstruction of the particle trajectory, machine learning can go a long way in taking the particle physics of the Large Hadron Collider to a new level [39,40]. The collimators need to be aligned in accordance with the configuration of the Large Hadron Collider. The alignment settings need to be monitored and updated in accordance with the requisite parameters. It needs to be noted at this point that the calorimeter data is divided in such a way that any alignment that is not in consonance with the set of given parameters can be done away with during the time of data cleansing [41,42]. This is important because the misalignments affect the performance of the calorimeter in the long run. The calorimeter hierarchy may also be pushed towards definite retraction for the optimization of performance. To set up a definite trajectory of angular accuracy, this optimization of performance proves to be a crucial step [43,44]. Different types of alignment methods are introduced and the optimal angle order is defined by making use of various machine learning techniques. In this alignment, techniques to maneuver the jaws of the calorimeter are developed until a clear and constant alignment in the spike arrangement is obtained. In the next steady-state, losses larger than the first ones may be witnessed [45,46]. There are a lot of factors responsible for these losses. Some of them include instability, mechanical vibrations, and unstable alignment [47,48]. Hence, there arises a need to analyze these losses which can be done effectively with the aid of machine learning techniques.

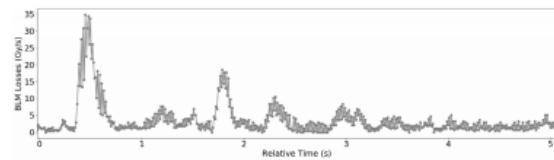


Fig. 2- (a) Graph showing alignment spikes when the collimator gets aligned according to the beam.

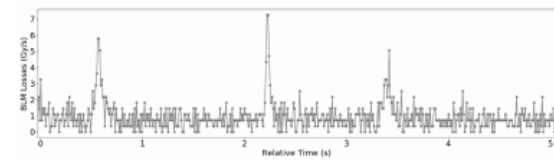


Fig. 2- (b) Graph showing non-alignment spikes when the collimator is not aligned according to the beam.

It is at this stage that the potential of machine learning techniques is recognized in treating the spike recognition problem as a classification problem. The machine learning models can be trained to effectively classify the sets of spike patterns observed in the losses [49]. For this, the data from various types of collimator alignment sets has been gathered. More than 6000 samples were collected for the study in which about 4000 spikes were positive and 3000 spikes were negative for which they were not perfectly aligned [50,51]. Structured data sets have been extracted from the collected data and the losses that have exceeded a specific threshold have been underlined. The technique of feature extraction may also be used to mine out features from this data set for deep analysis [52,53]. Soon after this, the most characteristic features and the degree of association between different pairs of variables has been observed using spearman correlation [54]. This process of feature selection enables us to perform our analysis on different types of machine learning models which is a precursor to determining the importance of each feature [55]. In the end, the features in an ascending order are obtained which are arranged by the degree of their importance. The forward selection algorithm has been used to zero in on the best features and also the best hyperparameters [56]. The forward selection algorithm does this by selecting one feature at a time and combining it with other features and finally extracts the best feature for future consideration. It is important to get some idea about the different types of features obtained as results of the study [57,58]. The first such feature is height. For the

calculation of height, the average of the steady state losses has been determined and subtracted from the peak value obtained for the spike [59,60]. The second important feature is that of spike decay. Another important feature is the determination of the normal distribution of various types of signals [61,62].

Using these value features, different machine learning models can be trained for classification. The machine learning models include support vector machines, decision trees, random forest models, logistic regression, and convolutional neural networks [63,64].

The synchronization of the calorimeter and beam alignment is necessary or else misalignment would result from inappropriate results. In addition to this, errors have been encountered in the detection of spike alignment which may affect the performance in the long run [65,66]. Post the detection of spikes, the performance of the model needs to be assessed. There is also need to analyze each subset in depth. In the end, the results need to be collected with the help of cross-validation on the training set [67,68]. There is also a need to divide the training set into various subsets for the validation of results. Repetition of this procedure n number of times enables us to partition the subset for the handling of splits [69]. In Fig 3, depiction of the simple techniques is given which is obtained from different models. The results are testimony to the fact that the support vector machine has ensured great precision [70]. The mean obtained in simple model technique has been similar to the mean obtained with the help of a support vector machine. The simple model also enables us to turn semi-automatic alignment into an automated one [71,72].

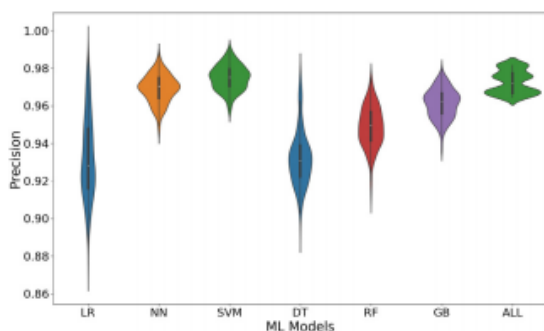


Fig. 3- Graph showing the precision distribution of various machine learning models.

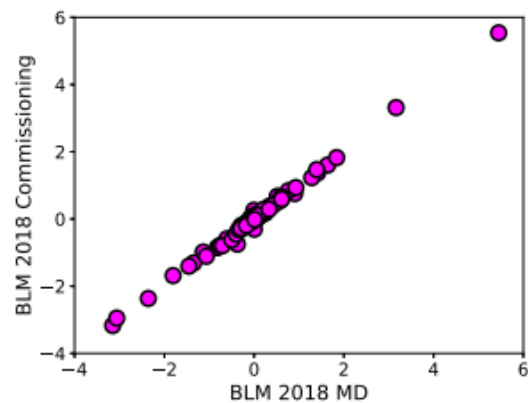


Fig. 4- Plot showing the correlation between median values in MD to the respective beam centers (2018 commissioning).

For the operation of the Large Hadron Collider in 2018, an automatic version of the software was used. Testing the alignment of the collimator is based on machine learning techniques which also enables us to collect data based on angular alignment. In Fig 4, a diagram is shown that is reproduced with the help of the Large Hadron Collider using the optics correction measurements that help in determining the stability of the collimator alignment [73,74]. The concern, however, is that it may take a lot of time to align the collimator in a required manner. This is where machine learning again comes to our aid and helps in decreasing the manual alignment time by about 70%. This is again depicted in Fig 5. At this stage, the research is still not equipped with fully automatic tools. The time requirement of alignment can be further reduced which even requires human intervention [75,76]. Full automation is not possible without using machine learning techniques. Such techniques can prove to be extremely active in interpreting the results and increasing the efficiency of the model [77,78].

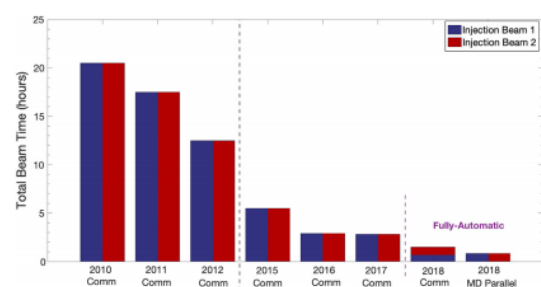


Fig. 5- Comparison graph between alignment time of collimators to the parallel alignment as per 2018 records.

4. How to minimize losses using machine learning

The Large Hadron Collider is an aggregate of numerous systems. Even the smallest of the losses aggregate into massive losses [79]. This makes it extremely difficult to build a model which not only identifies particles but also minimizes particle losses in the Large Hadron Collider. However, the conceiving of such a model would provide deep insights into the working of various systems of the Collider [80]. It needs to be noted at this point that the application of machine learning techniques is pivotal to the design and operation of such systems in the Large Hadron Collider. The intention has been to develop a system that would determine an operational set of parameters so that the intensity of the beam can be maximized while applying machine learning techniques [81,82]. This thing would be of great importance in the configuration of Large Hadron Collider and would help in identification of different correlations for comparative purposes. This would enable us to derive insights from experimental data with the help of particle trajectory simulation [83,84].

The main goal of this paper is to envisage a system that can predict the sequential set of operational parameters to maximize the intensity of the beam within a specific machine configuration [85]. This particular configuration may aid in the setup of the Large Hadron Collider and greatly help in computing correlations between various machine learning models. Finally, the main aim is to distinguish between various systems constructed from training data with inputs from particle trajectory graphs. The methodology utilized to construct the components is to rely on the huge sets of LHC data [86,87]. Such data sets are derived from several computational systems for constructing a data-based alternative model of the lifetime of the beam from the source to the end. There may be follow up with an optimization algorithm to compute the various parameters. It needs to be noted at this point that the supervised learning methodology has been used in the above case. The output of this system is the total trajectory of the beam and the inputs are the machine components including the tunes and magnetic parameters [88,89]. The data utilized in this case covers a series of repeated experiments. To simplify the feedback loop, the data is

taken from a representative set of various observations of large hadron collider experiments [90].

The first step was to train numerous types of ML models and compare them with the prevailing ones. Promising results have been obtained with the model called the Gradient Boosted Decision Tree model. After this, an alternative model was trained. This model could be tied up with many types of optimizers. This was followed by numerous types of experiments that yielded similar results [91,92]. The most effective optimizer has been used to mine out the required machine configuration from the model that we configured. In most of the cases, it was found that the input data deviated by quite some amount from the ideal. This is what was expected as an outcome, as the operation of the Large Hadron Collider relies largely on sticking to the same configuration settings in each cycle to avoid uncontrolled beam losses in various cases. Alternatively, a different model trained on various types of data represents operational machine learning parameters [93,94]. The limitation, however, is that it shows more deviation from prescribed machine learning parameters. The parameters that we considered were examined in detail with the help of customized controls in which multiple scans were done during different time intervals. The data sets gathered via different experiments were used to train and supplement the working model [95]. Several instabilities have also been witnessed in the beam trajectories that vary over some time, thereby, decreasing the performance of the model [96].

These instabilities are not considered by the present model which is a visible limitation of the current setup. Nevertheless, upon neglecting this aspect and limiting the study to the erstwhile naive, complete optimization query, the model seems to comply with the current optimal sectors of the vertical Q_v vs horizontal Q_h tune diagrams, as shown in Fig. 6. This model can be tuned to move towards the optimal regions, with maximum performance, obtaining a great improvement in the process [97,98]. This could be due to the reason that the maxima seen are very distant from the zone of operation. It has been concluded that the working model would have been sampled a relatively fewer number of times. The observation that the zone of operation fails to coincide with the maxima is due to the small gap between the tunes at the optimum

operating point [99]. As of now, it is difficult to ensure a low linear betatron coupling consistently. Thus, it becomes difficult to avoid accidental beam losses. In any case, it may be planned to migrate closer to the optimal operational zone in the next LHC cycle [100].

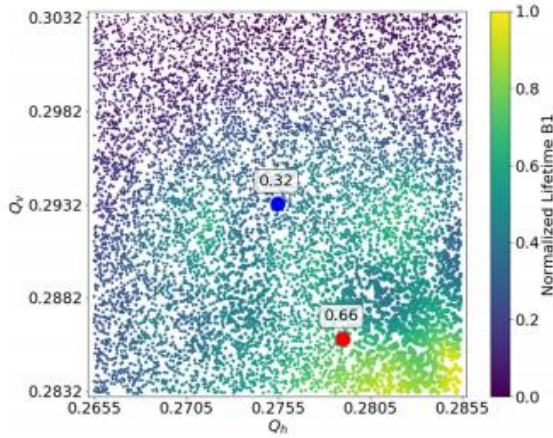


Fig. 6- Normalization plot of the lifetime of the beam versus LHC tune working point for a specific beam.

To treat the problem holistically, the above model will have to undergo improvisation by taking into account various types of instabilities, thereby framing a diverse optimization setup [101]. It needs to be noted at this point that this work is a nouveau approach to use data-driven ML techniques to conceive alternative models of the beam trajectories at the Large Hadron Collider. Although the models conceived are still not devoid of limitations, attempt is to make them generate optimal values that overlap with those derived from the actual readings [102].

5. How to detect collective beam instabilities

The problem with instability is that it can spoil the beam quality and also reduce its intensity. In this way, the luminosity of the Collider would decrease. For effective operation of the Large Hadron Collider, it is important to figure out such errors and find appropriate means of mitigating them [103]. One of the prime modes of mediation is the use of measurement devices like damper's observation boxes. The data which is gathered from the different monitors is collected by this box. This box also writes the buffer to the corresponding disk. Both manual and automatic procedures can be adopted but the automatic system is much more ideal for the situation [104,105]. The data collected shows different transfer beam positions

which cover different types of beam models. The data collected is almost of the order of 4 terabytes. To simplify the observations, the study has been restricted to the horizontal motion of beam one. However, it has to be acknowledged that the findings can be extended to other types of beams in the plane of operation [106].

To figure out the instabilities, the enormous amounts of data need to be ransacked which is a herculean task. This is where machine learning techniques like clustering are helpful in identifying useful beam oscillation patterns. The process of clustering rapidly helps us in classifying the various sets of triggers simultaneously. However, before performing any type of clustering, the removal of the false triggers from the data set needs to be ensured. After the filtering is done, the computational algorithms are run on different types of data sets [107]. This would greatly help in anomaly detection problems after the normal samples are collected. The process of filtering involves the use of the Fourier coefficient and related transformations. The complexity estimate is also used which helps in quantifying time series in the best possible manner. When the data set is gigantic, the principal component analysis can be used to reduce the number of dimensions and extract a set of features that can easily be worked upon [108]. The lowest possible dimensions in the case of principal component analysis can be found.

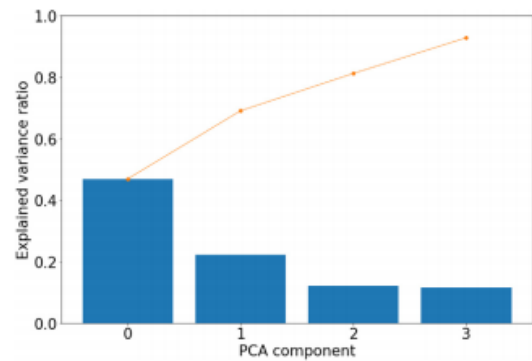


Fig. 7- Histogram displaying the variance ratio of various PCA components.

After this, the isolation forest algorithm is applied to the PCA space so that a similar sample can be isolated as shown in Fig 7. The isolation forest model splits the sample space repetitively by making use of hyperplanes. It is in this step that fewer iterations are used in anomalous rather than nominal points to make the distinction. It needs

to be noted at this point in time that there is also the possibility of manually classifying instabilities that need to be clubbed into the data set [109] as shown in Fig 8.

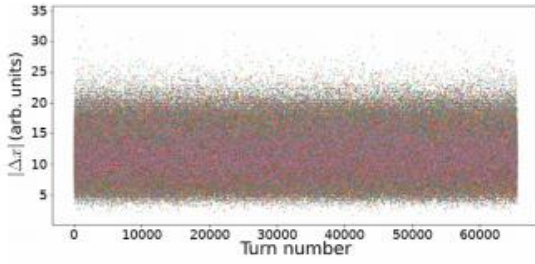


Fig. 8- (a) Plot showing the inlier according to the isolation forest model.

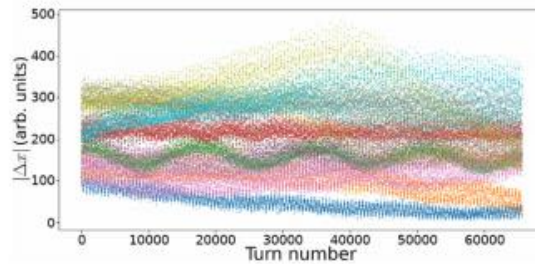


Fig. 8-(b) Plot showing the outlier according to the isolation forest model.

After the false triggers have been isolated and filtered out, work is done only on the anomalous samples from the entire data set. Different types of computational algorithms can be performed to mine out different results [110,111]. The model rests on the premises of a clustering algorithm with a univariate time series analysis. In this type of clustering, the distance metric criteria is required, and the distance metric that is used in this case is called dynamic time warping. The advantage of dynamic time warping is that it can compute the similarity between two-time series samples without getting affected by the local warping. This is preceded by the distance matrix to compute similarity in each group with respect to other groups. The distance metric obtained is passed on to the hierarchical clustering algorithm [112]. This is the algorithm that generates a dendrogram from the entire data set along with the signal as shown in Fig 9.

6. How to detect heating from pressure readings

It is a common fact that heating may be induced in grounding equipment due to high pressure. The

components that will lead to an increase in heating can be stated first. These include electron clouds, high-density variations, impedance and other types of heating that are induced by the electron beam [113]. The heating of the above components may lead to a large deviation from the result which may halt the performance of the machine during run time.

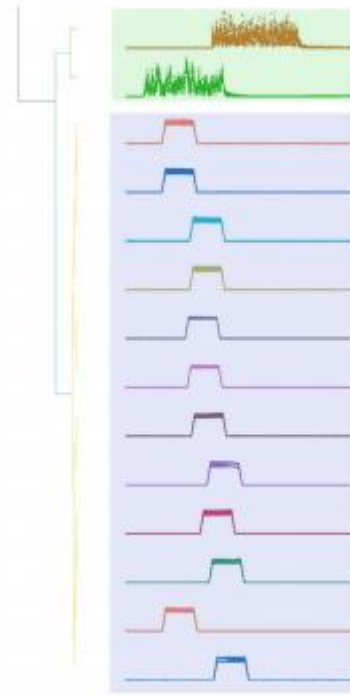


Fig. 9- Fig showing hierarchically clustered signals with highlighted beam offsets and instabilities.

Temperature probes have been used to monitor the heating induced by the beam. Optical fibers may also be used in some of the detectors so that the effect of heating is substantially reduced. However, it is not possible to cover the entire rings of the Large Hadron Collider by optical fibers and a little deviation from the result can be expected. The process of outgassing due to high temperature in a high vacuum environment is a point in the case. For instance, inside the injection protection device, the increase in temperature due to high pressure is particularly common. It is not difficult to tackle these anomalies as there is need to be ready for this [114]. The major problem that is difficult to tackle arises when the outgassing level becomes extremely abnormal and is observed in the beam gas ionization profile. The pressure may also increase to large amounts in some of the vacuum modules. This can be tackled by a modification in the architecture of the particular module. All the above

anomalies suggest that the smooth functioning of the Large Hadron Collider can be ensured by early detection of heating effects. Coming to vacuum monitoring, it is observed that this is relatively a systematic process when compared to temperature monitoring [115]. More than 1000 vacuum gauges are aligned to the circumference of the Large Hadron Collider. The task on hand is the analysis of the readings from the vacuum gauges one after another. Once an anomaly is detected, it needs to be rectified as quickly as possible so that there is no effect on the Large Hadron Collider trials. The challenge here is to quantify the vacuum reading into temperature values [116]. This is where machine learning could help in the quantification and investigation by automation of the pattern classification process. Machine learning can help in building an algorithm which automatically classifies the pressure reading for detection of heating spikes [117]. It is often observed (**Fig 10**) that there is an enormous increase in pressure induced by a vacuum gauge. It needs to be noted at this point that the sudden changes that occur in pressure readings do not coincide with bunch length variations. This indicates that it is the increase in temperature which has induced the process of outgassing [118].

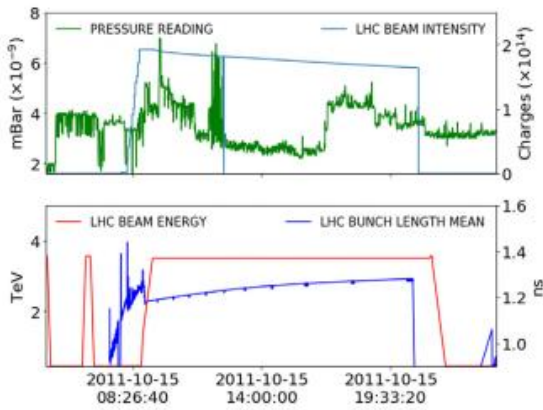


Fig. 10- Estimation of beam heating effects from different pressure measurements.

The aim here is to set up a classifier that decreases the time to detect anomalies in gauges. This suggests that classified should have a higher recall score. It has been found that the recall score shows the ratio of true positives to the total number of positives computed over a period [119]. About 1,000 readings have been collected and labelled appropriately leading to the creation of a data set with

more than 5000 steps for time series analysis of vacuum gauge reading. The next step has been to use the principal component analysis to reduce the dimensionality in which only a dozen features have been retained [120]. These features account for the variance that is observed in the dataset. This counters any type of information loss that may occur in the dataset. Two important steps are followed next which include the training of K-Nearest Neighbour classifier and multilayer perceptron [121]. The appropriate value of K needs to be determined using grid search. After this, the output needs to be passed via a logistic function for classification.

A 4-fold cross-validation technique has been used to evaluate the performance of a model. To calculate the percentage of samples, an additional technique called stratified splitting has been used [122]. The result of this splitting for maximization of recall for K nearest neighbor is shown in Fig 11. In the Fig 11, the digit sequence of red lines is observed which indicates overfitting for the training data set. This indicates that the model gives good results for the training set but gives poor results for other types of test data [123]. That sets the index to seven in the graph. As such, the model does not work to fit the training data and the value for the recall is obtained which is less than 1. In Fig 11, the index is set to 0 and in this case, the recall score between point 86 and point 1 has been opted for. This represents a substantial improvement when compared with the nearest neighbor case K.

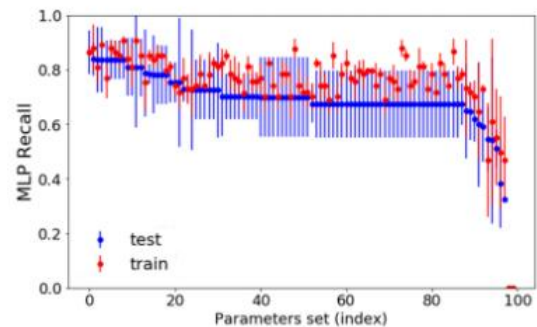
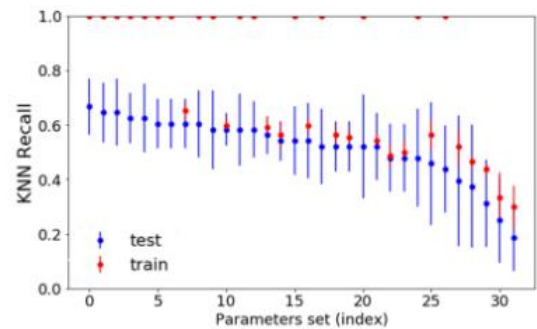


Fig. 11- Graph showing the recall scores obtained using the KNN classifier and the MLP classifier.

7. Experimental Results

The learning curves obtained for the support vector machine have been shown in Fig 12. The true positives, false positives, and false negatives have been depicted by blue, green and red lines respectively. The results obtained from anomaly detection have been shown in Fig 13. In this Fig, the number of samples have been plotted against true positives, false positives, false negatives and true negatives. In order of occurrence, SVM, DBSCAN, LOF, Binary OR and post-processing after DBSCAN have been shown by blue, green, red, black and violet colours respectively. True positives show the anomaly that is correctly detected while true negatives show the normal point that is correctly detected. The likeness between DBSCAN results obtained after processing and ground truth is depicted in Fig 14. In Fig 15, a 3D visualization concerning anomalies for ground truth and results of post-processing after DBSCAN have been depicted. Finally, in Fig 16.1 to Fig 16.10, the results obtained after measuring the top tagging efficiencies and statistical uncertainties for various data samples are shown.

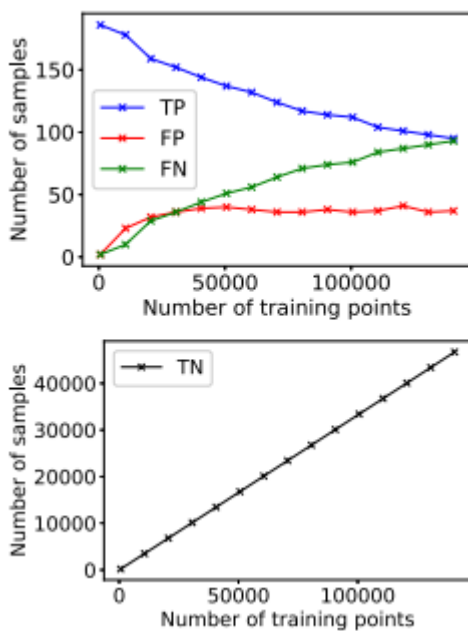


Fig. 12- Fig showing SVM training curves including true positives, false positives, true negatives, and false negatives.

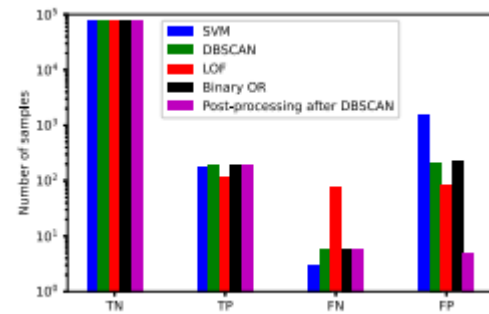


Fig. 13- Fig showing anomaly detection results obtained by various ML techniques.

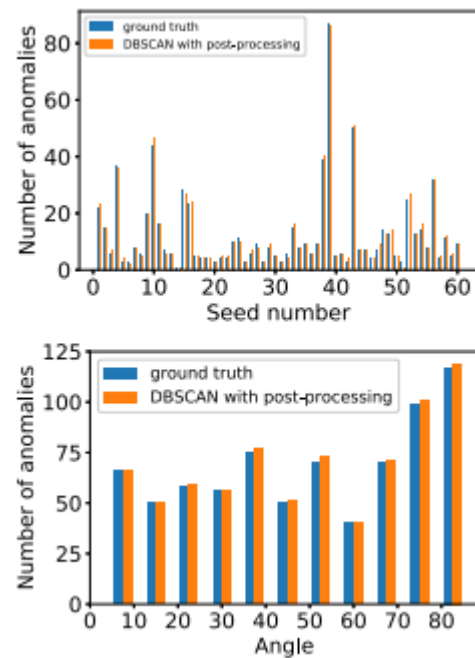


Fig. 14- Plot showing the number of anomalies (seed and angle) by comparing ground truth with post-processing DBSCAN results.

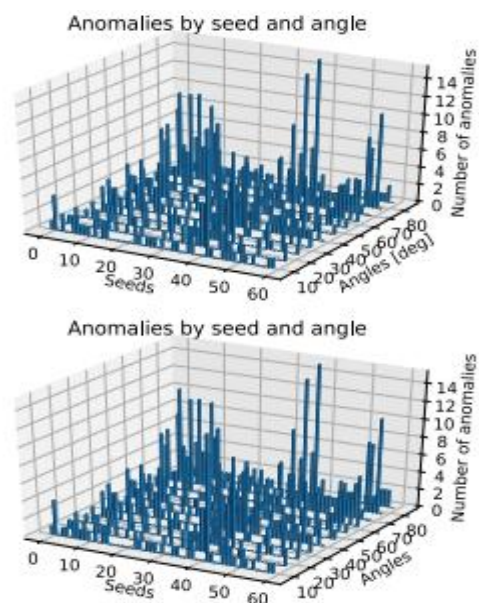


Fig. 15- A 3D plot of the number of anomalies (seed and angle) obtained by comparing ground truth with post-processing DBSCAN results.

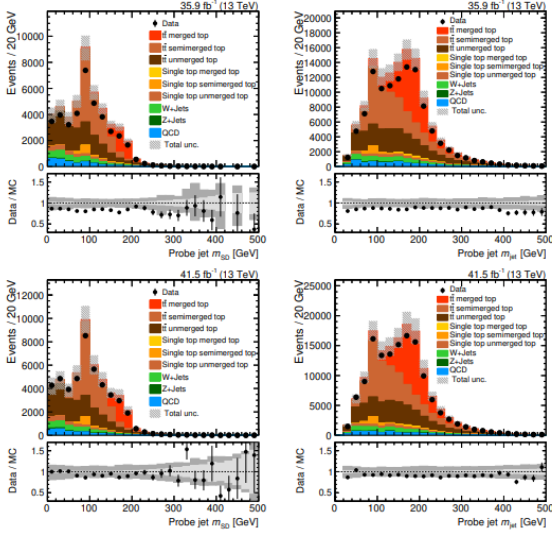


Fig 16.1- The graphs show the ratio of events observed to that predicted by simulation for the first set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

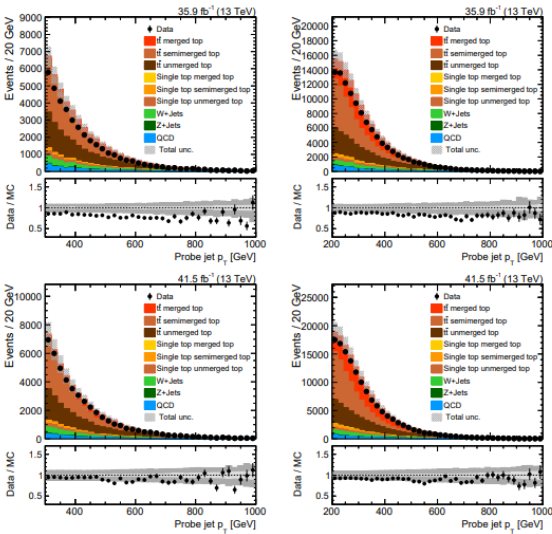


Fig 16.2- The graphs show the ratio of events observed to that predicted by simulation for the second set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

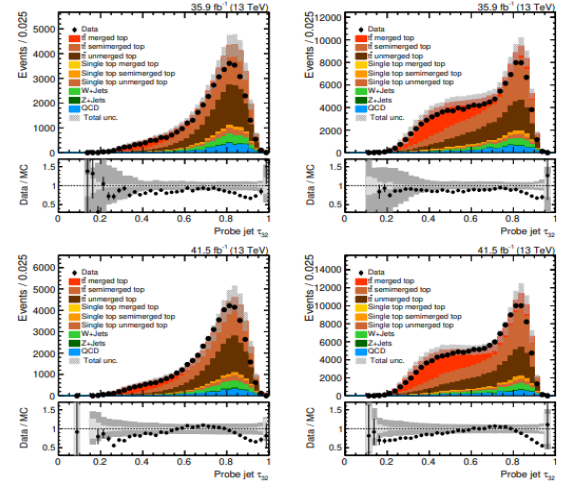


Fig 16.3- The graphs show the ratio of events observed to that predicted by simulation for the third set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

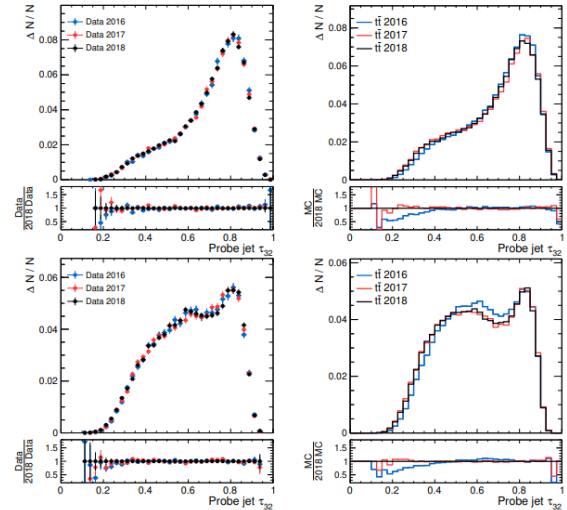


Fig 16.4- The normalized distributions of the relative difference for data collected to that prediction via simulation.

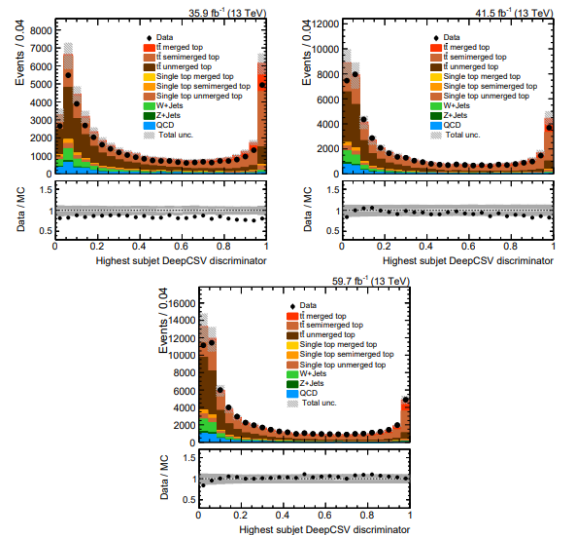


Fig 16.5- The graphs show the ratio of events observed to that predicted by, simulation for the fifth set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs the statistical uncertainty and total uncertainty are displayed separately.

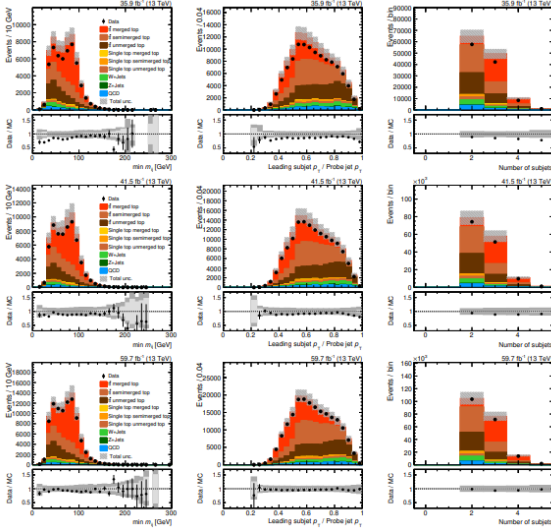


Fig 16.6- The graphs show the ratio of events observed to that predicted by simulation for the sixth set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

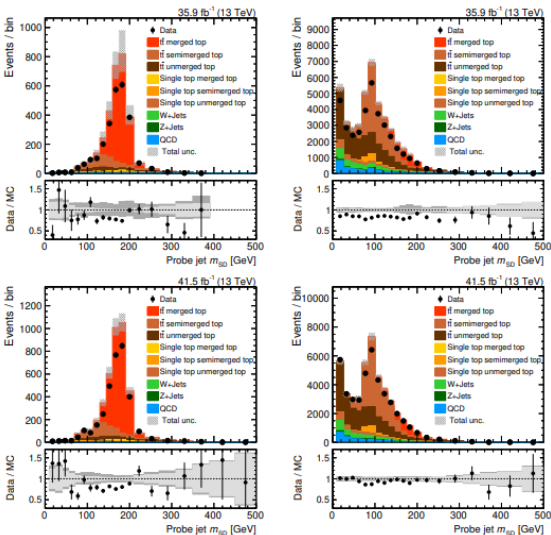


Fig 16.7- The graphs show the ratio of events observed to that predicted by simulation for the seventh set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

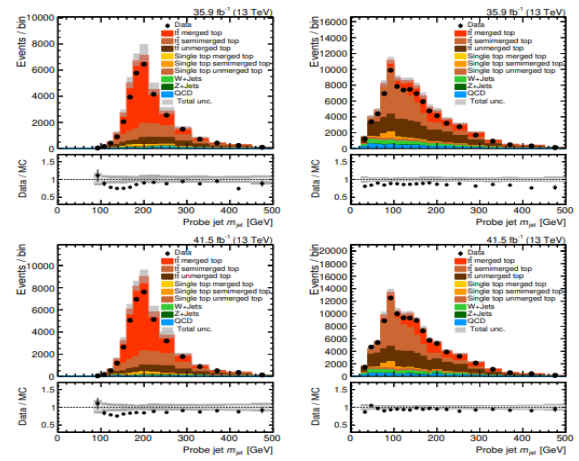


Fig 16.8- The graphs show the ratio of events observed to that predicted by simulation for the eighth set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

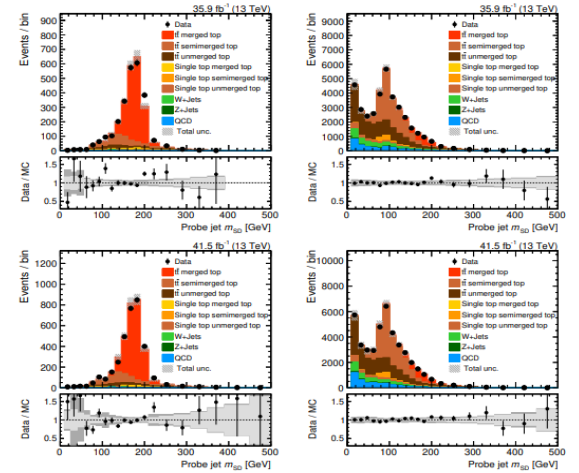


Fig 16.9- The graphs show the ratio of events observed to that predicted by simulation for the ninth set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

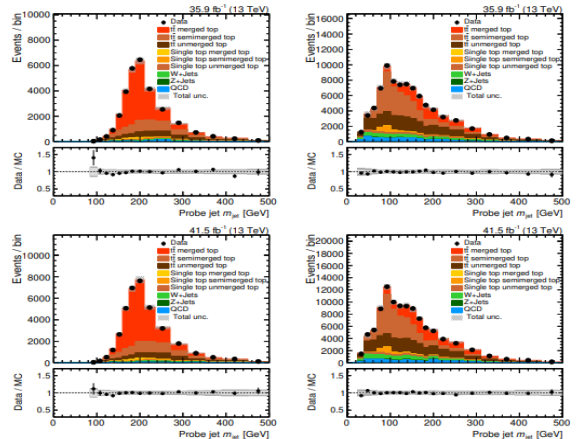


Fig 16.10- The graphs show the ratio of events observed to that predicted by simulation for the tenth set of samples. In the first two graphs, the total uncertainty is shown by the shaded area and, in the second two graphs, the statistical uncertainty and total uncertainty are displayed separately.

8. Conclusion

In this paper, the researchers have strived to present a detailed overview of the machine learning techniques which can be applied to the beam dynamics at the Large Hadron Collider. Both supervised learning and unsupervised learning find great applications in particle identification, particle trajectory tracking, and minimization of beam losses. Promising results have been obtained after being able to distinguish between actual beam loss spikes and those obtained erroneously. All this became possible by application of machine learning techniques.

In near future, the evolution of particle physics, collimation systems and beam dynamics would be studied in consonance with machine learning techniques. The research that is proposed in this paper would serve as a guiding document for prospective models that aim to study the particle physics at the Large Hadron Collider from the lens of machine learning methodologies.

References

- [1] M. Giovannozzi, F.F. Van der Veken, Description of the luminosity evolution for the CERN LHC including dynamic aperture effects. Part II: application to Run 1 data, Nucl. Instrum. Methods Phys. Res. A 908 (2018) 1–9, arXiv:1806.03059.
- [2] F.F. Van der Veken, M. Giovannozzi, Scaling laws for the time dependence of luminosity in hadron circular accelerators based on simple models of dynamic aperture evolution, in: Proc. 61st ICFA Advanced Beam Dynamics Workshop on High-Intensity and High-Brightness Hadron Beams, HB2018, Daejeon, Korea, 2018, pp. 260–265.
- [3] M.M. Breunig, H.-P. Kriegel, R.T. Ng, J. Sander, LOF: Identifying density-based local outliers, in: Proc. of the 2000 ACM SIGMOD International Conference on Management of Data, Dallas, USA, 2000, pp. 93–104.
- [4] M. Giovannozzi, F.F. Van der Veken, Description of the luminosity evolution for the CERN LHC including dynamic aperture effects. Part I: the model, Nucl. Instrum. Methods Phys. Res. A 905 (2018) 171–179, [Erratum: Nucl. Instrum. Methods Phys. Res. A A927, 471(2019)], arXiv:1806.03058.
- [5] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in: Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining, KDD'96, Portland, OR, USA, 1996, pp. 226–231.
- [6] G. Batista, X. Wang, E. Keogh, A complexity-invariant distance measure for time series, in: Proceedings of the 11th SIAM International Conference on Data Mining, SDM 2011, 2011, pp. 699–710.
- [7] G. Azzopardi, A. Muscat, S. Redaelli, B. Salvachua, G. Valentino, Operational results of LHC collimator alignment using machine learning, in: Proc. 10th Int. Particle Accelerator Conf., IPAC'19, Melbourne, Australia, 2019, pp. 1208–1211.
- [8] G. Azzopardi, S. Redaelli, B. Salvachua, A. Muscat, G. Valentino, Automatic beam loss threshold selection for LHC collimator alignment, in: Proc. 17th Int. Conf. on Accelerator and Large Experimental Physics Control Systems, ICALEPCS'19, New York, USA, 2019.
- [9] B. Salvant, O. Aberle, G. Arduini, R. Aßmann, V. Baglin, M. Barnes, W. Bartmann, P. Baudrenghien, O. Berrig, C. Bracco, E. Bravin, G. Bregliozzi, R. Bruce, A. Bertarelli, F. Carra, G. Cattenoz, F. Caspers, S. Claudet, H. Day, M. Garlasche, L. Gentini, B. Goddard, A. Grudiev, B. Henrist, R. Jones, O. Kononenko, G. Lanza, L. Lari, T. Mastoridis, V. Mertens, E. Métral, N. Mounet, J. Muller, A. Nosych, J. Nougaret, S. Persichelli, A. Piguet, S. Redaelli, F. Roncarolo, G. Rumolo, B. Salvachua, M. Sapinski, R. Schmidt, E. Shaposhnikova, L. Tavian, M. Timmins, J. Uythoven, A. Vidal, J. Wenninger, D. Wollmann, M. Zerlauth, P. Fassnacht, S. Jakobsen, M. Deile, Update on beam induced RF heating in the LHC, in: Proc. 4th International Particle Accelerator Conf., IPAC'13, Shanghai, China, 2013, pp. 1646–1648.
- [10] E. Fol, J.M. Coello de Portugal, R. Tomás, Unsupervised machine learning for detection of faulty beam position monitors, in: Proc. 10th Int. Particle Accelerator Conf., IPAC'19, Melbourne, Australia, 2019, pp. 2668–2671.
- [11] R. Giusti, G.E.A.P.A. Batista, An empirical comparison of dissimilarity measures for time series classification, in: 2013 Brazilian Conference on Intelligent Systems, 2013, pp. 82–88.
- [12] E. Fol, J.M. Coello de Portugal, G. Franchetti, R. Tomás, Application of machine learning to beam diagnostics, in: Proc. 39th Int. Free Electron Laser Conf., FEL'19, Hamburg, Germany, 2019.
- [13] A. Edelen, C. Mayes, D. Bowring, D. Ratner, A. Adelman, R. Ischebeck, J. Snuverink, I. Agapov, R. Kammering, J. Edelen, I. Bazarov, G. Valentino, J. Wenninger, Opportunities in machine learning for particle accelerators, 2018, arXiv:1811.03172.
- [14] E. Bozoki, A. Friedman, Neural networks and orbit control in accelerators, in: Proc. 4th European Particle Accelerator Conf.,

- EPAC'94, London, England, 1994, pp. 1589–1592, [1589, 1994].
- [15] C.M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag, Berlin, Heidelberg, 2006
- [16] ATLAS Collaboration, CMS Collaboration and LHC Higgs Combination Group, Procedure for the LHC Higgs boson search combination in Summer 2011, CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, (2011).
- [17] S. Schmitt, TUnfold: an algorithm for correcting migration effects in high energy physics, JINST 7 (2012) T10003, doi: 10.1088/1748-0221/7/10/T10003, arXiv:1205.6201
- [18] T. Sjöstrand and M. van Zijl, A multiple interaction model for the event structure in hadron collisions, Phys. Rev. D 36 (1987) 2019, doi: 10.1103/PhysRevD.36.2019.
- [19] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, N-Jettiness: An Inclusive Event Shape to Veto Jets, Phys. Rev. Lett. 105 (2010) 092002, doi: 10.1103/PhysRevLett.105.092002, arXiv:1004.2489.
- [20] J. Alwall, H. S., F. Krauss et al., Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions, Eur. Phys. J. C 53 (2008) 473, doi: 10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [21] A. H. Hoang, S. Mantry, A. Pathak et al., Extracting a short distance top mass with light grooming, Phys. Rev. D 100 (2019) 074021, doi: 10.1103/PhysRevD.100.074021, arXiv:1708.02586.
- [22] J. Alwall, R. Frederix, S. Frixione et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations, JHEP 07 (2014) 079, doi: 10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [23] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma et al., Lightgbm: A highly efficient gradient boosting decision tree, in *Advances in Neural Information Processing Systems* 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan et al., eds., pp. 3146–3154, Curran Associates, Inc., (2017)
- [24] M. Ester, H.-P. Kriegel, J. Sander and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD'96)*, Portland (OR), USA, (1996) 226
- [25] E. Fol, J. M. Coello de Portugal, G. Franchetti and R. Tomás, Application of Machine Learning to Beam Diagnostics, in *Proc. 39th Int. Free Electron Laser Conf. (FEL'19)*, Hamburg, Germany, (2019).
- [26] F. T. Liu, K. M. Ting and Z.-H. Zhou, Isolation forest, in *Proc. 8th IEEE Int. Conf. on Data Mining (ICDM'08)*, IEEE Computer Society, (2008) 413.
- [27] G. Azzopardi et al., Automatic Beam Loss Threshold Selection for LHC Collimator Alignment, in *Proc. 17th Int. Conf. on Accelerator and Large Experimental Physics Control Systems (ICALEPCS'19)*, New York (NY), USA, (2019).
- [28] R. W. Aßmann et al., Expected performance and beam-based optimization of the LHC collimation system, in *Proc. 9th European Particle Accelerator Conf. (EPAC'04)*, Lucerne, Switzerland, (2004) 1825.
- [29] F. Ravotti et al., Radiation monitoring in mixed environments at CERN, from the IRRAD6 facility to the LHC experiments, IEEE Trans. Nucl. Sci. 54 (2007) 1170.
- [30] F. Ravotti et al., Conception of an integrated sensor for the radiation monitoring of the CMS experiment at the large hadron collider, IEEE Trans. Nucl. Sci. 51 (2004) 3642.
- [31] J Albrecht, V V Gligorov, G Raven, and S Tolk. Performance of the LHCb High Level Trigger in 2012. Jan 2014. arXiv:1310.8544.
- [32] R Antunes-Nobrega et al. LHCb reoptimized detector design and performance: Technical Design Report. Technical report, CERN, Geneva, 2003.
- [33] M. Pojer, M. Solfaroli Camillocci, the fraction of events which can be labelled as uninteresting Schmidt, R. Nevertheless, and S. Wagner. Machine Availability at the Large Hadron Collider. Conf.Proc., C1205201:3566–3568, 2012.
- [34] M Williams, V. V. Gligorov, C Thomas, H Dijkstra, J Nardulli, and P Spradlin. The HLT2 Topological Lines. Technical Report LHCb-PUB-2011-002, CERN, Geneva, Jan 2011.
- [35] Kell, G.S. Density, thermal expansivity, and compressibility of liquid water from 0. deg. to 150. deg.. Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale. J. Chem. Eng. Data 1975, 20, 97.
- [36] Kadam, G.P.; Mishra, H. Dissipative properties of hot and dense hadronic matter in an excluded-volume hadron resonance gas model. Phys. Rev. C 2015, 92, 035203
- [37] ALICE Collaboration. Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV. Phys. Rev. C 2019, 99, 024906.
- [39] Zyla, P.A.; Barnett, R.M.; Beringer, J.; Dahl, O.; Dwyer, D.A.; Groom, D.E.; Lin, C.-J.; Lugovsky, K.S.; Pianori, E.; Robinson, D.J.; et al. (Particle Data Group), Review of Particle Physics. Prog. Theor. Exp. Phys. 2020, 2020, 083C01.
- [38] Khuntia, A.; Tiwari, S.K.; Sharma, P.; Sahoo, R.; Nayak, T.K. Effect of Hagedorn states on isothermal compressibility of hadronic matter formed in heavy-ion collisions: From NICA to LHC energies. Phys. Rev. C 2019, 100, 014910.
- [39] Tawfik, N.A.; Abou-Salem, L.I.; Shalaby, A.G.; Hanafy, M.; Sorin, A.; Rogachevsky, O.; Scheinast, W. Particle production and chemical freezeout from the hybrid UrQMD approach at NICA energies. Eur. Phys. J. A 2016, 52, 324
- [40] Deppman, A. Properties of hadronic systems according to the nonextensive self-consistent thermodynamics. J. Phys. G Nucl. Part. Phys. 2014, 41, 055108

- [41] Cleymans, J.; Worku, D. The Tsallis distribution in proton-proton collisions at $\sqrt{s} = 0.9$ TeV at the LHC. *J. Phys. G* 2012, 39, 025006.
- [42] Chatrchyan, S.; Khachatryan, V.; Sirunyan, A.M.; Tumasyan, A.; Adam, W.; Aguilo, E.; Bergauer, T.; Dragicevic, M.; Erö, J.; Fabjan, C.; et al. Study of the inclusive production of charged pions, kaons, and protons in pp collisions at $\sqrt{s} = 0.9, 2.76$, and 7 TeV. *Eur. Phys. J. C* 2012, 72, 2164
- [43] Aamodt, K.; Abel, N.; Abeysekara, U.; Quintana, A.A.; Abramyan, A.; Adamova, D.; Aggarwal, M.M.; Rinella, G.A.; Agocs, A.G.; Salazar, S.A.; et al. Production of pions, kaons and protons in pp collisions at $\sqrt{s} = 900$ GeV with ALICE at the LHC. *Eur. Phys. J. C* 2011, 71, 1655.
- [44] Bhattacharyya, T.; Cleymans, J.; Marques, L.; Mogliacci, S.; Paradza, M. W. On the precise determination of the Tsallis parameters in proton-proton collisions at LHC energies. *J. Phys. G* 2018, 45, 055001
- [45] Ortiz, A.; Paí'c, G.; Cuautle, E. Mid-rapidity charged hadron transverse sphericity in pp collisions simulated with Pythia. *Nucl. Phys. A* 2015, 941, 78.
- [46] Werner, K.; Karpenko, I.; Pierog, T.; Bleicher, M.; Mikhailov, K. Event-by-event simulation of the three-dimensional hydrodynamic evolution from flux tube initial conditions in ultrarelativistic heavy ion collisions. *Phys. Rev. C* 2010, 82, 044904.
- [47] Nayak, R.; Pal, S.; Dash, S. Effect of rope hadronization on strangeness enhancement in p-p collisions at LHC energies. *Phys. Rev. D* 2019, 100, 074023
- [48] V. Sharma, S. Rai, A. Dev, "A Comprehensive Study of Artificial Neural Networks", *International Journal of Advanced Research in Computer Science and Software Engineering*, ISSN 2277128X, Volume 2, Issue 10, October 2012
- [49] J. M. Keller, M. R. Gray, J. A. Givens Jr., "A Fuzzy K-Nearest Neighbor Algorithm", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-15, No. 4, August 1985
- [50] P. Harrington, "Machine Learning in Action", Manning Publications Co., Shelter Island, New York, ISBN 9781617290183, 2012
- [51] S. B. Hiregoudar, K. Manjunath, K. S. Patil, "A Survey: Research Summary on Neural Networks", *International Journal of Research in Engineering and Technology*, ISSN: 2319 1163, Volume 03, Special Issue 03, pages 385-389, May, 2014
- [52] D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical Study", *Journal of Artificial Intelligence Research*, 11, Pages 169- 198, 1999
- [53] L. P. Kaelbling, M. L. Littman, A. W. Moore, "Reinforcement Learning: A Survey", *Journal of Artificial Intelligence Research*, 4, Page 237-285, 1996
- [54] X. Zhu, A. B. Goldberg, "Introduction to Semi - Supervised Learning", *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2009, Vol. 3, No. 1, Pages 1-130
- [55] R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", *Machine Learning*, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992
- [56] M. Andrecut, "Parallel GPU Implementation of Iterative PCA Algorithms", *Institute of Biocomplexity and Informatics*, University of Calgary, Canada, 2008
- [57] S. S. Shwartz, Y. Singer, N. Srebro, "Pegasos: Primal Estimated sub - Gradient Solver for SVM", *Proceedings of the 24th International Conference on Machine Learning*, Corvallis, OR, 2007
- [58] K. Alsabati, S. Ranaka, V. Singh, "An efficient k-means clustering algorithm", *Electrical Engineering and Computer Science*, 1997
- [59] L. Rokach, O. Maimon, "Top - Down Induction of Decision Trees Classifiers - A Survey", *IEEE Transactions on Systems*,
- [60] Carvalho, T.P.; Soares, F.A.A.M.N.; Vita, R.; da Francisco, P.R.; Basto, J.P.; Alcalá, S.G.S. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* 2019, 1, 1-12.
- [61] "ITU-T Focus Group on Machine Learning for Future Networks including 5G." [Online]. Available: <https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx>
- [62] D. Woods and T. J. Naughton, "Optical computing: Photonic neural networks," *Nature Physics*, vol. 8, no. 4, pp. 257-259, July 2012.
- [63] D. Brunner, M. Soriano, C. Mirasso, and I. Fischer, "High speed, high performance all-optical information processing utilizing nonlinear optical transients," in *The European Conference on Lasers and ElectroOptics*, May 2013.
- [64] H. Akaike, *Akaike's Information Criterion*, M. Lovric, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
- [65] A. P. Vela, M. Ruiz, and L. Velasco, "Applying Data Visualization for Failure Localization," in *Optical Fiber Communication Conference*, Mar. 2018.
- [66] G. Choudhury, D. Lynch, G. Thakur, and S. Tse, "Two Use Cases of Machine Learning for SDN-Enabled IP/Optical Networks: Traffic Matrix Prediction and Optical Path Performance Prediction," to appear in *IEEE/OSA Journal of Optical Communications and Networking*, 2018.
- [67] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153-1176, Oct. 2015.
- [68] T. T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification using machine learning," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 56-76, 4th Q 2008.
- [69] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, June 2014.

- [70] P D Stachour and B M Thuraishingham Design of LDV A multilevel secure relational database management system, IEEE Trans. Knowledge and Data Eng., Volume 2, Issue 2, 190 - 209, 1990
- [71] Adankon, M. and Cheriet, M. Genetic algorithm-based training for semi-supervised SVM, Neural Computing and Applications , Volume 19(8), 1197-1206, 2010
- [72] P. Donmez and J. G. Carbonell, Proactive learning: cost-sensitive active learning with multiple imperfect oracles. In Proceedings of the 17th ACM conference on Information and knowledge management, 619–628. ACM, 2008
- [73] Zhiyang J, Weiwei L, Wei G, Youming X (2012) Research on web spam detection based on support vector machine. In: Communication Systems and Network Technologies (CSNT), 2012 International Conference on, pp 517–520, DOI 10.1109/csnt.2012.117
- [74] Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Philip SY (2008) Top 10 algorithms in data mining. Knowledge and Information Systems 14(1):1–37
- [75] Yuchun T, Krasser S, Yuanchen H, Weilai Y, Alperovitch D (2008) Support vector machines and random forests modeling for spam senders behavior analysis. In: Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE, pp 1–5, DOI 10.1109/glocom.2008.ecp.419
- [76] Rodriguez JJ, Kuncheva LI, Alonso CJ (2006) Rotation forest: A new classifier ensemble method. Pattern Analysis and Machine Intelligence, IEEE Transactions on 28(10):1619–1630
- [77] Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The annals of statistics 28(2):337–407
- [78] Cortes C, Vapnik V (1995) Support-vector networks. Machine Learning 20(3):273–297, DOI 10.1007/bf00994018, URL <http://dx.doi.org/10.1007/BF00994018>
- [79] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.
- [80] Karystinos, George N., and Dimitrios A. Pados. "On overfitting, generalization, and randomly expanded training sets." IEEE Transactions on Neural Networks 11.5 (2000): 1050-1057
- [81] Kato, Nei, et al. "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective." IEEE wireless communications 24.3 (2017): 146-153
- [82] J. M. Keller, M. R. Gray, J. A. Givens Jr., "A Fuzzy KNearest Neighbor Algorithm", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4, August 1985
- [83] S. S. Shwartz, Y. Singer, N. Srebro, "Pegasos: Primal Estimated sub - Gradient Solver for SVM", Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, 2007
- [84] R. S. Sutton, "Introduction: The Challenge of Reinforcement Learning", Machine Learning, 8, Page 225-227, Kluwer Academic Publishers, Boston, 1992
- [85] S. B. Hiregoudar, K. Manjunath, K. S. Patil, "A Survey: Research Summary on Neural Networks", International Journal of Research in Engineering and Technology, ISSN: 2319 1163, Volume 03, Special Issue 03, pages 385-389, May, 2014
- [86] M. Andrecut, "Parallel GPU Implementation of Iterative PCA Algorithms", Institute of Biocomplexity and Informatics, University of Calgary, Canada, 2008
- [87] D. Opitz, R. Maclin, "Popular Ensemble Methods: An Empirical Study", Journal of Artificial Intelligence Research, 11, Pages 169- 198, 1999
- [88] X. Zhu, A. B. Goldberg, "Introduction to Semi – Supervised Learning", Synthesis Lectures on Artificial Intelligence and Machine Learning, 2009, Vol. 3, No. 1, Pages 1- 130
- [89] K. Alsabati, S. Ranaka, V. Singh, "An efficient k-means clustering algorithm", Electrical Engineering and Computer Science, 1997
- [90] L. P. Kaelbling, M. L. Littman, A. W. Moore, "Reinforcement Learning: A Survey", Journal of Artificial Intelligence Research, 4, Page 237-285, 1996
- [91] S. S. Shwartz, Y. Singer, N. Srebro, "Pegasos: Primal Estimated sub - Gradient Solver for SVM", Proceedings of the 24th International Conference on Machine Learning, Corvallis, OR, 2007
- [92] M. Bowles, "Machine Learning in Python: Essential Techniques for Predictive Analytics", John Wiley & Sons Inc., ISBN: 978-1-118- 96174-2
- [93] C. Sutton, T. Hobson, J. Geddes, and R. Caruana, "Data diff: Interpretable, executable summaries of changes in distributions for data wrangling," in Proc. of the 24th ACM SIGKDD. ACM, 2018, pp. 2279–2288.
- [94] C. Hill, R. Bellamy, T. Erickson, and M. Burnett, "Trials and tribulations of developers of intelligent systems: A field study," in Visual Languages and Human-Centric Computing (VL/HCC), 2016 IEEE Symposium on. IEEE, 2016, pp. 162–170.
- [95] D. Bohus, S. Andrist, and M. Jalobeanu, "Rapid development of multimodal interactive systems: a demonstration of platform for situated intelligence," in Proc. of the 19th ACM International Conference on Multimodal Interaction. ACM, 2017, pp. 493–494.
- [96] L. Bottou, "Stochastic gradient descent tricks" in Neural Networks: Tricks Trade, Berlin, Germany:Springer, pp. 421-436, 2012
- [97] Mohammed Javeed Zaki. Scalable algorithms for association mining. IEEE transactions on knowledge and data engineering, 12(3):372–390, 2000.
- [98] Peng Wei, Yufeng Li, Zhen Zhang, Tao Hu, Ziyong Li, and Diyang Liu. An optimization method for intrusion detection

- classification model based on deep belief network. *IEEE Access*, 7:87593–87605, 2019.
- [99] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, et al. Constrained k-means clustering with background knowledge. In *Icml*, volume 1, pages 577–584, 2001.
- [100] Wei Wang, Jiong Yang, Richard Muntz, et al. Sting: A statistical information grid approach to spatial data mining. In *Vldb*, volume 97, pages 186–195, 1997.
- [101] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [102] Iqbal H Sarker and ASM Kayes. Abc-ruleminer: User behavioral rule-based machine learning method for context-aware intelligent services. *Journal of Network and Computer Applications*, page 102762, 2020.
- [103] Hae-Sang Park and Chi-Hyuck Jun. A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, 36(2):3336–3341, 2009.
- [104] Youssoufa Mohamadou, Aminou Halidou, and Pascal Tiam Kapen. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of covid-19. *Applied Intelligence*, 50(11):3913–3925, 2020.
- [105] Mohssen Mohammed, Muhammad Badruddin Khan, and Eihab Bashier Mohammed Bashier. *Machine learning: algorithms and applications*. Crc Press, 2016.
- [106] Lior Rokach. A survey of clustering algorithms. In *Data Mining and Knowledge Discovery Handbook*, pages 269–298. Springer, 2010.
- [107] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [108] Andrew McCallum. Information extraction: Distilling structured data from unstructured text. *Queue*, 3(9):48–57, 2005.
- [109] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.
- [110] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.
- [111] Bing Liu Wynne Hsu Yiming Ma. Integrating classification and association rule mining. In *Proceedings of the fourth international conference on knowledge discovery and data mining*, 1998.
- [112] Huan Liu and Hiroshi Motoda. *Feature extraction, construction and selection: A data mining perspective*, volume 453. Springer Science & Business Media, 1998.
- [113] Maurice Houtsma and Arun Swami. Set-oriented mining for association rules in relational databases. In *Data Engineering, 1995. Proceedings of the Eleventh International Conference on*, pages 25–33. IEEE, 1995.
- [114] Saskia Le Cessie and Johannes C Van Houwelingen. Ridge estimators in logistic regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 41(1):191–201, 1992.
- [115] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.
- [116] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [117] Robert C Holte. Very simple classification rules perform well on most commonly used datasets. *Machine learning*, 11(1):63–90, 1993.56. S. Sathya Keerthi, Shirish Krishnaji Shevade, Chiranjib Bhattacharyya, and Karuturi Radha Krishna Murthy. Improvements to platt’s smo algorithm for svm classifier design. *Neural computation*, 13(3):637–649, 2001.
- [118] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [119] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.
- [120] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [121] Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40, 1975.
- [122] Mihael Ankerst, Markus M Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: ordering points to identify the clustering structure. *ACM Sigmod record*, 28(2):49–60, 1999.
- [123] Amitabha Das, Wee-Keong Ng, and Yew-Kwong Woon. Rapid association rule mining. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 474–481. ACM, 2001.