EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

September 1, 2023

Summer Student Report

Particle Identification in the Muon Forward Tracker using Machine Learning

Author:
Theresa Reisch
theresa.reisch@cern.ch

Supervisors:
Luca Micheletti
Stefano Politanò

Abstract

In this work, the particle identification capabilities of the ALICE Muon Forward Tracker (MFT) are explored. The goal is to detect (anti-)nuclei at forward rapidity ($-3.6 < \eta < -2.4$) for the first time. This measurement is crucial to improve rapidity dependent uncertainties in nucleosynthesis models. The study presented in this note is performed on the data collected by the ALICE experiment during RUN 3. A Machine learning algorithm is developed to optimise the separation of nuclei and Minimum Ionising Particles (MIP) in the MFT using anomaly detection. The resulting statistical significance for nuclei is > 30 which means that the MFT can be used to detect (anti-)nuclei at forward rapidity. In combination with the high charge-reconstruction efficiency of the MFT for $p_{\mathrm{T}} < 2\,\mathrm{GeV}$, anti-nuclei could be distinguished.

Contents

1 Introduction

Light anti-nuclei like anti-^{3}He can originate from high energy cosmic ray collisions with the interstellar medium or from annihilation of unknown dark matter particles in our Galaxy. At the Large Hadron Collider, anti-nuclei and their interaction with ordinary matter can be studied with high precision. The ALICE experiment has already studied the production of light nuclei at mid-rapidity ($|\eta| < 0.5$) [1] indicating that anti-^{3}He nuclei can be used to study cosmic-ray interactions and dark matter annihilation. These results are used for nucleosynthesis models for dark matter searches in cosmic rays [2–5]. But one of the main uncertainties in these models is the rapidity dependence of the nucleosynthesis process.

In this study, possibilities for the first measurements of light nuclei in the forward direction with the ALICE detector are investigated. In the following Chapter, Section 2, the detector setup used is described. In Section 3, the Machine Learning methods used are explained and in Section 4 the results for the particle identification are presented. Lastly, Section 5 concludes with the results and gives an outlook on the future.

2 ALICE Experiment

The ALICE experiment [6] is optimised to study heavy-ion collisions at a centre-of-mass energy up to 5.36 TeV per nucleon pair. Proton-proton collisions provide an important baseline for heavy ion collisions and allow to test predictions of perturbative QCD calculations. The production of light nuclei was also studied in pp collisions [1]. The data for this analysis originates from the first year of LHC Run 3 in pp collisions with a centre-of-mass energy of $\sqrt{s} = 13.6\,\text{TeV}$. For Run 3, a major upgrade took place for ALICE, introducing new components like the Muon Forward Tracker 2.2. Although the ALICE detector consists of multiple different subsystems, only the forward parts of the detector are relevant for this analysis.

2.1 Muon Spectrometer

In the forward direction, the relevant detector for this study is the Muon Spectrometer [7] which can be seen in Fig. 1. During the last upgrade for Run 3, the Muon Forward Tracker (MFT) [8] was newly installed. The Muon Spectrometer consists of the MFT, the Muon CHambers (MCH) and the Muon IDentifier (MID). The MFT is closest to the interaction point and can track all charged particles in the forward direction ($-3.6 < \eta < -2.4$). It is followed by the front absorber which is supposed to stop hadrons. The MCH is located after the absorber and can track all charged particles that have passed the first absorber, i.e. mostly muons. The tracks reconstructed in the MFT and MCH can be connected and a χ^2 is calculated characterising how well the two tracks in the subsystems match. The MID is located behind a second absorber and works as a triggering detector. Because of the two absorbers before, muons are identified with a high purity in this subsystem. This means the to identify muon tracks in the MFT, the track has to pass both absorbers and has to be connected to a track in the MCH with a small χ^2. A track satisfying these conditions is defined as a global track.

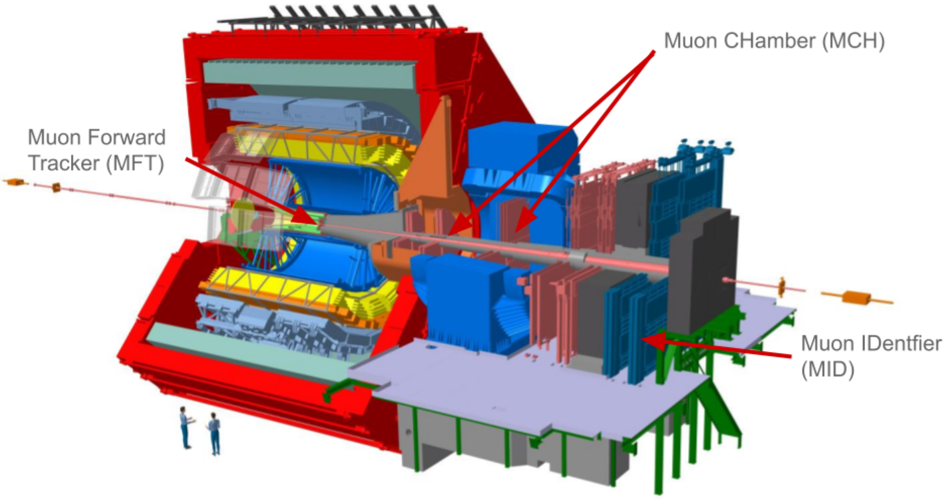


Figure 1: The ALICE detector with the Muon Spectrometer. The Muon Forward Tracker is the closest to the interaction point. It is then followed by the front absorber and the Muon Chambers. After this, there is a second absorber and the Muon Identifier. ©ALICE.

2.2 Muon Forward Tracker

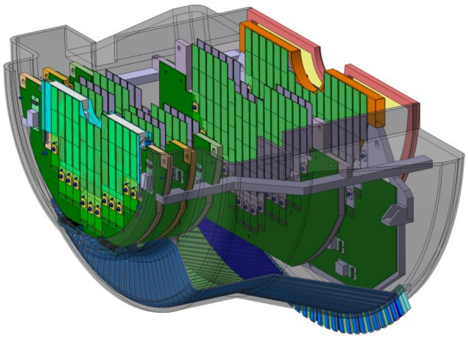


Figure 2: The Muon Forward Tracker of ALICE. It consists of 5 disks of silicon pixel sensors positioned along the beam axis in the forward direction. ©ALICE.

The Muon Forward Tracker measures tracks of charged particles in forward direction with the pseudorapidity coverage of η of -3.6 to -2.4. The MFT detector surrounds the vacuum beam pipe and is positioned in front of the Muon Spectrometer. As can be seen in Fig. 2, it consists of 5 double-faced disks positioned along the beam axis at $z = \{$-460, -493, -531, -687, -786$\}$ mm from the interaction point. The disks consist of ALPIDE silicon pixel sensors [9] with a binary read out.
At midrapidity, a magnetic field of 0.5 T is used allowing for the reconstruction of the momentum which is used for particle identification. The position of the MFT lays outside of the magnet which means that no sufficiently good momentum resolution is provided and the particle identification cannot be done based on the momentum like at midrapidity. Nevertheless, the residual curvature can be used to distinguish the charge of the particle. The charge-reconstruction efficiency (shown in Fig. 3) is high according to Run 3 MC simulations, which means that discrimination between matter and anti-matter is possible.
When charged particles hit the silicon pixel sensors of the MFT, they create a signal. The signal in the MFT is produced by the charge diffusion and collection in the silicon layer which results in a cluster where a pixel can either be activated or not. For an analog read-out, the cluster size would be proportional to the energy loss -$\langle \mathrm{d}E/\mathrm{d}x \rangle$ of the particle hitting the detector. For the binary read-out of the MFT,

this is expected as well but needs to be checked.
The Bethe-Bloch formula

$$-\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = \frac{4\pi e^4 z^2 n Z}{m v^2} \left[\ln\left(\frac{2mv^2}{I}\right) - \ln\left(\frac{1}{\beta^2}\right) - \frac{\beta^2}{2} \right] \tag{1}$$

describes this energy loss by a charged particle traversing a layer of material. It depends on the velocity v, mass m, charge number z of the incoming particle and electron density n and mean excitation energy I of the material. Consequently, according to the Bethe-Bloch formula the energy loss is proportional to the charge number of the incoming particle squared. Due to the Lorentz boost, the particles hitting the MFT behaves like a Minimum Ionising Particle (MIP) and can be described by the MIP region of the Bethe Bloch equation. In conclusion, particles with a higher charge number have a higher energy loss and should have a bigger cluster. That is the reason why the particle identification possibilities in the forward direction with the MFT are explored.

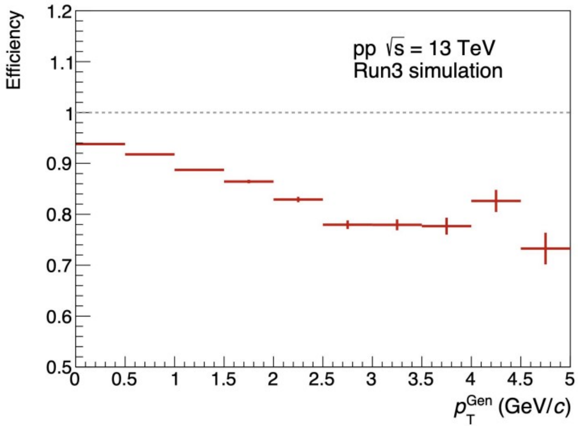


Figure 3: Charge reconstruction efficiency depended on the p_{T}.

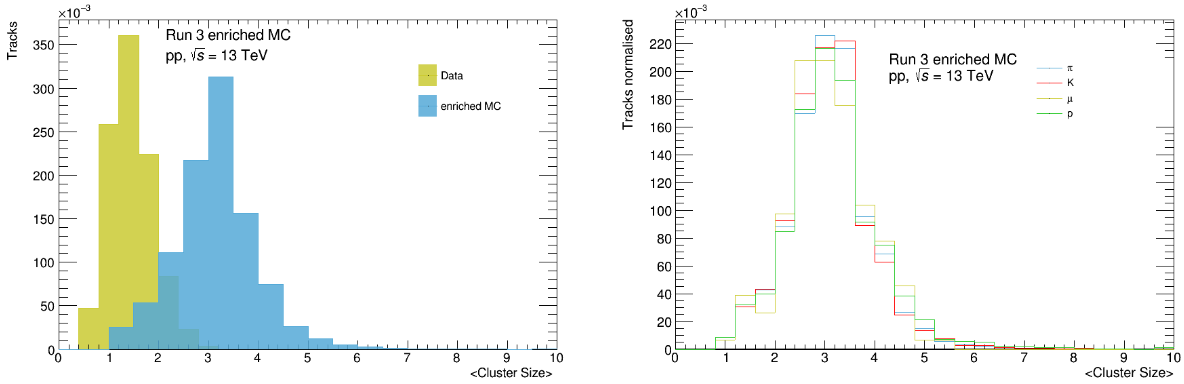
3 Machine Learning Methods

The aim of this work is to investigate whether Machine Learning tools can improve the signal of nuclei at forward rapidity. In order to do that, an algorithm able to distinguish between nuclei and Minimum Ionising Particles (MIPs) is needed.The machine learning models used for that purpose are introduced here.

3.1 Training Samples

A Machine Learning classification algorithm requires training, which is the process in which the algorithm ”learns” patterns and correlation of the input data in order to optimise the separation between different classes. The two approaches for the training are supervised and unsupervised learning. In supervised learning, a classification algorithm needs to be trained on all possible classes whereas in unsupervised learning, it is enough to have one of the classes as a ”normal” class and the other one as an ”anomalous” one. The anomaly detection principle is further explained in Section 3.3.

Usually, classification algorithms are supervised learning algorithms like Boosted Decision Trees [10]. For the supervised learning, MC simulations are suitable because training data for all classes exist. The MC simulation used in this analysis is a PYTHIA 8 [11] simulation with injected nuclei, which means in every event, one (anti-)nuclei is forced at generation level. The detector response and the transport of the generated particles is simulated using GEANT 4 [12]. In the case of the ALPIDE response and the cluster size, the MC simulation currently deviates from the data, as can be seen in Fig. 4a. In data, the MIPs (muons) have a smaller average cluster size than in the MC simulation. Therefore, at the current state, the MC cannot be used for training the machine learning model. Nevertheless, the MC can be useful for closure tests and in the MC, the signal can be enriched in order to have a larger data sample. As an alternative, the machine learning algorithm can be trained on data directly. In data, muons are a proxy for all MIP because the distributions of the average cluster size are similar in the MC simulation, as Fig. 4b shows. The muons can be filtered in data with a high purity by checking for global tracks that pass through all absorbers and have a low χ^2 for the matching of tracks between the MFT and MCH. In order to perform supervised learning, a training data sample of nuclei would be necessary as well, but filtering nuclei is not yet possible. As a conclusion, unsupervised learning and anomaly detection is performed (see Section 3.3).



(a) Comparison of the average cluster size of MIP in the enriched MC and data. The MC distribution is not compatible with the data.

(b) Comparison of the average cluster size of kaons, pions, protons and muons in MC. The distributions are all well compatible with each other.

Figure 4: Characterisation of the training samples.

3.2 Neural Networks

Artificial neural networks are a fundamental concept in machine learning, inspired by the structure of the human brain. Neural networks have revolutionised modern physics by improving methods of data analysis, reconstruction and calibration [13, 14]. They consist of layers of connected nodes, each performing simple computations on incoming data and passing the results to subsequent layers. Through the training, neural networks learn to recognise complex patterns and relationships within data. This training involves adjusting the weights and biases of the connections between neurons to minimise the difference between the network's predictions and the actual target values, usually described by the loss function.

In general, all machine learning models have hyperparameters that are set before the learning process started and are not changed in the training by the algorithm. In the case of a neural networks, those hyperparameters are the number of layers and the number of nodes per layers as well as the loss function, learning rate and the activation function. Therefore, the hyperparameter optimisation is an important step in using a machine learning algorithm.

3.3 AutoEncoders

AutoEncoders (AE) [15] are a type of artificial neural network used for unsupervised learning. An AutoEncoder consists of an encoder and a decoder with a minimal dimension bottleneck between them, as illustrated in Fig. 5. During the training, the encoder compresses the input data into a lower-dimensional latent space, while the decoder reconstructs the original data from this representation.

It can be used for anomaly detection which is the identification of anomalous data which deviate significantly from the ”normal” data. For that, the AE is trained on normal data and learns the normal pattern. Consequently, it performs well on recreating the normal data. If being applied on anomalous data, the

AE performs worse on recreating that data. By using the reconstruction error, deviations that exceed a certain threshold can be flagged as anomalies.

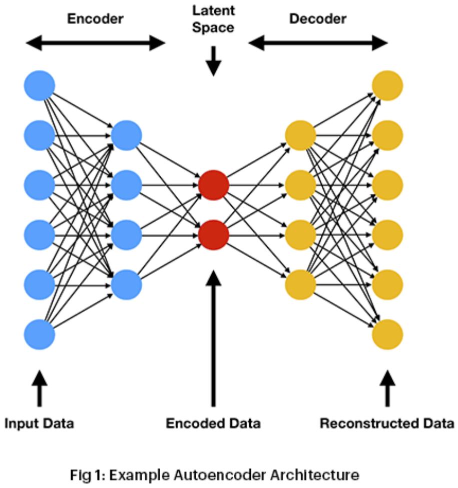
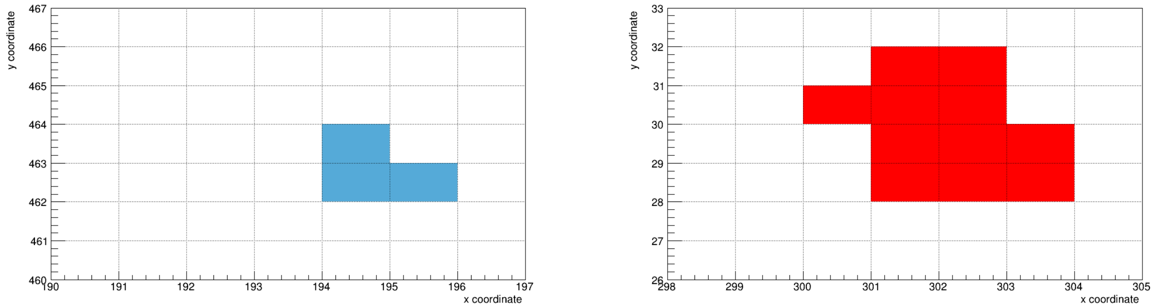


Figure 5: Architecture of an AutoEncoder consisting of an encoder, the latent space and decoder.

4 Particle Identification in the MFT

4.1 Linear Selection

Production of nuclei in forward direction is limited in contrast to MIP, that is the reason why MC simulation with injected nuclei is used for this preliminary study. As already explained in Section 2.2, the cluster size is related to the energy loss which is related to the charge number Z via the Bethe-Bloch formula (1). Because the MIPs (kaons, pions, protons, electrons and muons) all have charge number $z = 1$ and the light nuclei (^{3}He) have $z = 2$, the energy deposition should be four times larger for the nuclei. In Fig. 6, examples for a typical cluster of a MIP and nuclei are shown.



(a) MIP cluster.

(b) Nuclei cluster.

Figure 6: Examples of clusters created by MIP or nuclei in the MFT.

In the enriched MC, this can be observed, as shown in Fig. 7. The separation allows to define a selection above which everything is marked as a (anti-)nuclei. With this selection, the efficiency (ε), purity (P) and significance (sign.) can be calculated, characterising the performance of the selection:

$$\varepsilon = \frac{\text{true nuclei above threshold}}{\text{all true nuclei}},$$

$$P = \frac{\text{true nuclei above threshold}}{\text{all particles above threshold}},$$
$$\text{sign.} = \frac{\text{true nuclei above threshold}}{\sqrt{\text{all particles above threshold}}}.$$

The values for these variables obtained for different threshold values are displayed in Fig. 8. The maximum of the significance which is the best trade-off between efficiency and purity is the chosen working point for the comparison. Although the efficiency and purity are obtained using the enriched MC, the significance can be scaled to a realistic nuclei-to-MIP ratio $R = 2 : 10^7$ using the number of events for Run 2 $N_{\text{events}} = 492 \cdot 10^6$ [16] as reference and average number of particles per event $\langle N_{\text{particles}} \rangle = 14.58$ [16] according to

$$N_{\text{nuclei}} = N_{\text{events}} \cdot \langle N_{\text{particles}} \rangle \cdot R \cdot \varepsilon, \tag{2}$$
$$N_{\text{nuclei}} + N_{\text{MIP}} = N_{\text{nuclei}} / P. \tag{3}$$

With that, a maximal signal significance of 33 can be reached, as listed in Tab. 1, but it is still a tentative calculation.

significance	efficiency	purity
33	0.86	0.88

Table 1: Efficiency and purity for the optimal working point maximising the signal significance. The significance is scaled based on realistic numbers for MIP/nuclei ratio and total number of particles.

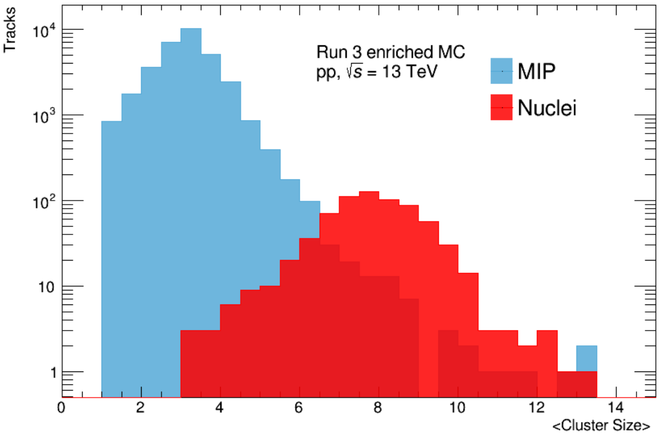


Figure 7: Comparison of the average cluster size for MIP and nuclei in the enriched MC sample. The distributions are well separable.

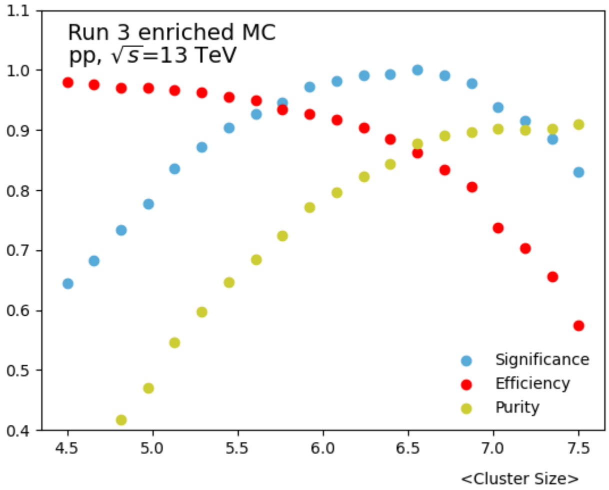


Figure 8: Efficiency, purity and significance for the nuclei in the enriched MC for different selections on the average cluster size. The significance is normalised to the maximum. The optimal working point is defined at the maximal significance.

4.2 AutoEncoder Selection

As already mentioned in Section 3.1 and 3.3, unsupervised learning with an AutoEncoder can deal with the problems presented. There are multiple AutoEncoder implementations available and in the tests, PyTorch [17] worked better than the competing Keras [18], as the comparison of the efficiencies and purities for different selections in Fig 9 shows.

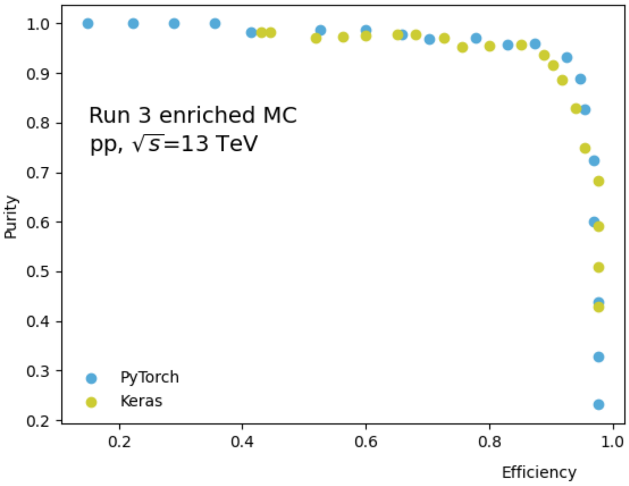


Figure 9: Comparison of Keras and PyTorch implementation of an AE with the same hyperparameters on the same dataset.

4.2.1 *Hyperparameter Optimization*

AutoEncoders are deep and complex machine learning algorithms that need hyperparameter optimization. For the hyperparameter optimisation, Optuna [19] was used which is a framework for efficient and automatic hyperparameter optimisation. The goal of the optimisation is to either minimise or maximise a defined objective function like the validation loss for the best performance of the algorithm. While doing

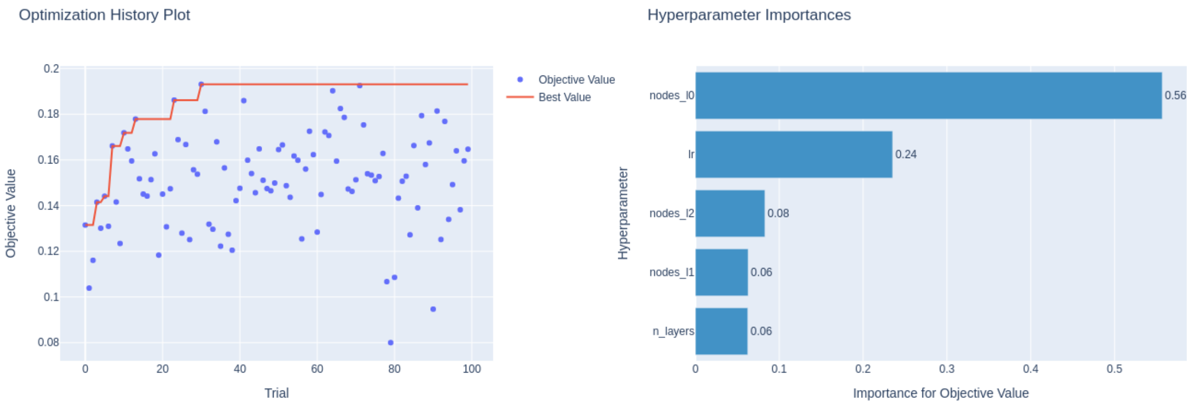
so, Optuna speeds up the optimisation by exploiting the previous iteration.
In the case of the anomaly detection, the chosen objective function is the difference between the mean of the validation loss for a normal and anomalous validation sample which is to be maximised. With that, it is secured that the model is not simply overfitting and minimising the validation loss, but is optimal for the anomaly detection. In a first step, the loss function, activation function, learning rate and architecture of the AE was optimised. Quickly, it became clear that the Mean Absolute Error (MAE) function

$$\mathrm{MAE} = \sum_{i=1}^{n} \left| y_{\mathrm{input}} - y_{\mathrm{reconstructed}} \right| \tag{4}$$

was the best loss function and the rectified linear unit (ReLU) function

$$\mathrm{ReLU}(x) = \max\left(0, x\right) \tag{5}$$

best suited for the activation. The further activation was done only considering the architecture consisting of the number of layers and nodes per layers and the learning rate. In the process of the optimisation, the objective value is calculated every time and the configuration with the (in this case) highest value is chosen. The history of the optimisation is visualised in Fig. 10a. In Fig. 10b, the importance of the individual hyperparameters are displayed showing clearly that the number of nodes of the first layer and the learning rate are the most important ones.



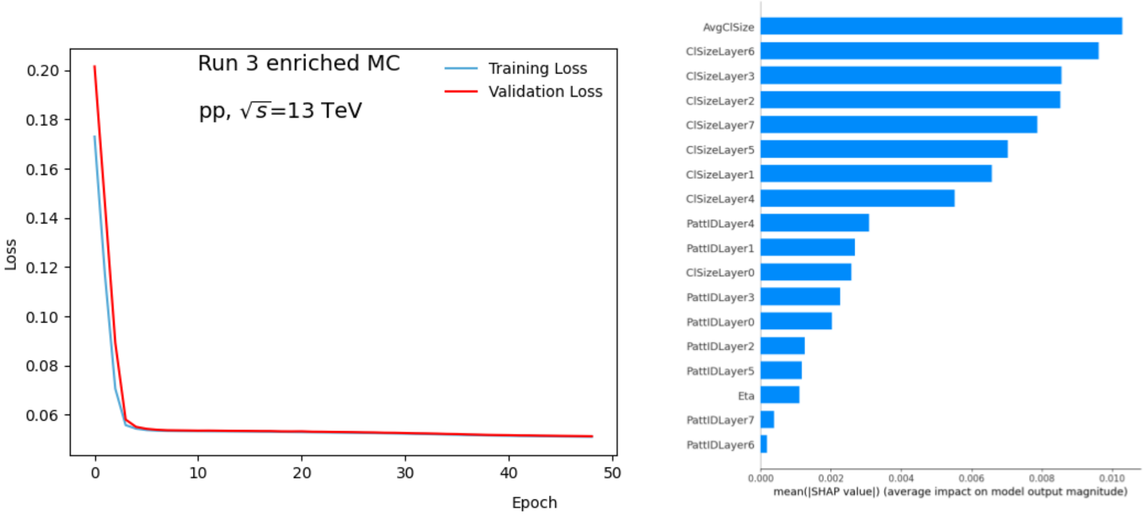
(a) Optimisation history of Optuna. (b) Hyperparameter importance of the optimisation.

Figure 10: Hyperparameter Optimisation with Optuna.

4.2.2 *Results with enriched MC*

The AutoEncoder is trained on a fraction (80%) of MIP (either MC MIP or data muons) which is considered the normal data for the anomaly detection. The performance of the AE can then be measured as the discrepancy of the AE output and original input. This reconstruction performance should be better on the trained MIP than on the nuclei anomalous data which can be used to define a threshold for the separation of nuclei and MIP.
Firstly, the AE is trained and tested purely on the enriched MC. The loss values for the training and validation sample over the epochs of the training is shown in Fig. 11a. The validation loss is always slightly higher than the training loss because the AE has not learnt the validation set but a bigger deviation would hint to the problem of overfitting where the model just learns the training set by heart and does not recognise the pattern. In Fig. 11b, the feature importance determined by SHAP [20] is shown. Clearly, the average cluster size has the highest impact as a feature, but the others features are taken into account as well. When removing the average cluster size as an input feature, the performance stays the same since

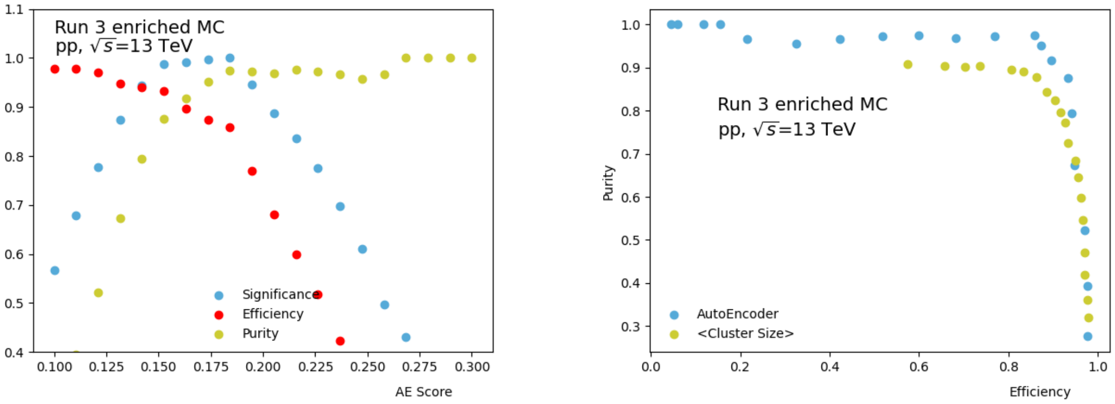
the AE is able to average the cluster size by itself. In conclusion, this leads to higher efficiency and purity when selecting on the AE score than the linear selection on the average cluster, as can be seen in Fig. 12. The signal significance quoted in Tab. 2 also exceeds the significance with the linear selection. The AE score distribution when testing on the enriched MC sample is shown in Fig. 13 and clearly shows two separable distributions.



(a) Loss for the training and validation sample over the epochs of the training.

(b) Summary of input feature importance during the training.

Figure 11: Training on the enriched MC.



(a) Efficiency, purity and significance for the nuclei for different selections on the AE score. The significance is normalised to the maximum.

(b) Comparison of efficiency and purity for the nuclei in the enriched MC for different selections on the AE score and average cluster size.

Figure 12: Results for training and testing on the enriched MC.

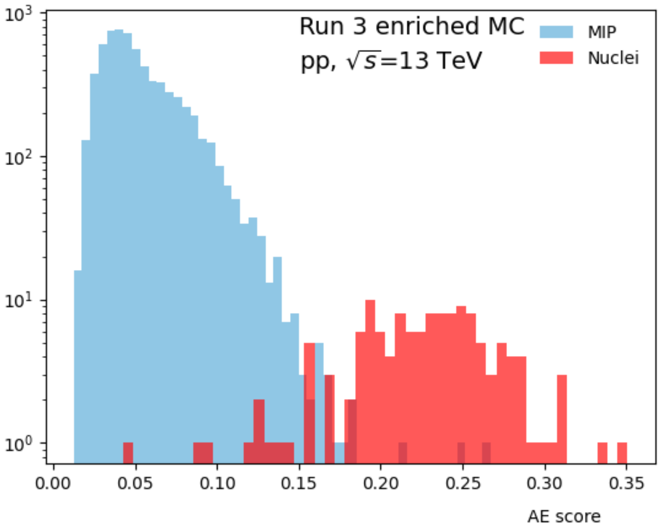


Figure 13: AE score for testing on MIP and nuclei for the enriched MC.

significance	efficiency	purity
35.2	0.93	0.93

Table 2: Efficiency and purity for the optimal working point maximising the signal significance. The significance is scaled based on realistic numbers for MIP/nuclei ratio and total number of particles.

4.2.3 Results with Data

For training on data, particles in the MFT are selected if they have a global track (i.e. signal in MFT, MCH and MID) and have a low $\chi^2(< 10)$ for the matching of the tracks in the MFT and MCH.

The AE was then tested on the enriched MC simulations and a test muon data set. The result for the AE can be seen in Fig. 14. The MC nuclei and MIP distribution are again well separable and the data muons are as expected in the MIP region of the distribution. From this, the signal significance, efficiency and purity can be calculated, as shown in Fig. 15a and listed in Tab. 3 for the optimal working point. The AE still outperforms the linear selection on the average cluster size, as shown in Fig. 15b and the significance is compatible to training and testing on the enriched MC. Nevertheless, these results are still preliminary because the current MC cannot exactly replicate the cluster size for the MIPs and perhaps the nuclei.

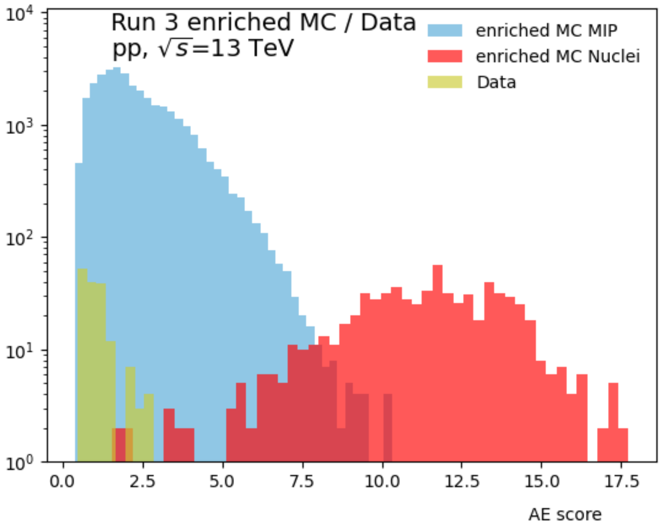
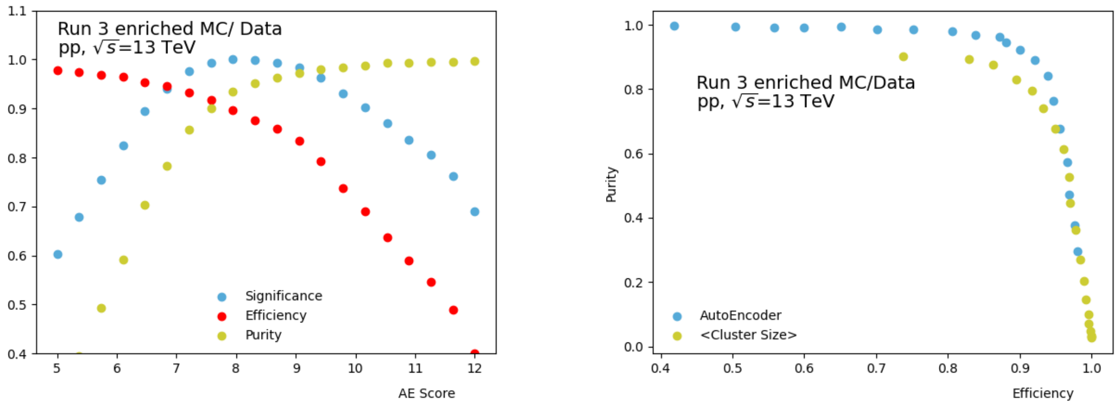


Figure 14: AE score for training on data and testing on data and MIP and nuclei from the enriched MC.

significance	efficiency	purity
34.7	0.90	0.93

Table 3: Efficiency and purity for the optimal working point maximising the signal significance. The significance is scaled based on realistic numbers for MIP/nuclei ratio and total number of particles.



(a) Efficiency, purity and significance for the nuclei for different selections on the AE score. The significance is normalised to the maximum.

(b) Comparison of efficiency and purity for the nuclei for different selections on the AE score and average cluster size.

Figure 15: Results for training on Data and testing on the enriched MC and Data.

5 Conclusion and Outlook

As described in the previous section, the MFT is able to identify nuclei with a significance of > 30 and efficiencies and purities > 0.85. Those numbers are corrected from the MC with enriched nuclei to realistic MIP/nuclei ratio and can give a realistic estimation.
The usage of Machine Learning with the optimised AutoEncoders taking multiple input variables can enhance the signal and outperform the linear selection on one of the input features. Although the performance is still preliminary because the testing can only be done using the MC simulations which are currently not compatible with data in terms of the cluster size of MIP. A next step would be to repeat the study with improved MC as soon as it is available.
In combination with the high charge-reconstruction efficiency, matter and anti-matter can be distinguished well, opening up the possibilities to measure anti-nuclei at forward rapidity in ALICE. This kind of measurements are crucial for understanding anti-nuclei in our Galaxy and studying Dark Matter and of interest for the whole high energy physics community.

Acknowledgements

I would like to thank my supervisors Luca Micheletti and Stefano Politanò for their invaluable guidance, unwavering support, and exceptional kindness throughout the entirety of this project. I would also like to thank the ALICE collaboration for welcoming me as part of the CERN Summer Student Programme 2023. Lastly, I would like to thank the organisers of this Summer Student Programme and all the nice people I have met here for this exceptional summer.

References

[1] **ALICE** Collaboration, "Measurement of anti-^{3}He nuclei absorption in matter and impact on their propagation in the Galaxy", *Nat. Phys.* **19** (2023) .

[2] A. Ibarra and S. Wild, "Prospects of antideuteron detection from dark matter annihilations or decays at ams-02 and gaps", *JCAP* **2013** no. 02, (Feb, 2013) . `https://dx.doi.org/10.1088/1475-7516/2013/02/021`.

[3] E. Carlson *et al.*, "Antihelium from dark matter", *Phys. Rev. D* **89** (Apr, 2014) . `https://link.aps.org/doi/10.1103/PhysRevD.89.076005`.

[4] M. W. Winkler and T. Linden, "Dark matter annihilation can produce a detectable antihelium flux through $\overline{\Lambda}_b$ decays", *Phys. Rev. Lett.* **126** (Mar, 2021) . `https://link.aps.org/doi/10.1103/PhysRevLett.126.101101`.

[5] M. Korsmeier, F. Donato, and N. Fornengo, "Prospects to verify a possible dark matter hint in cosmic antiprotons with antideuterons and antihelium", *Phys. Rev. D* **97** (May, 2018) . `https://link.aps.org/doi/10.1103/PhysRevD.97.103011`.

[6] **ALICE** Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) .

[7] **ALICE** Collaboration, *ALICE dimuon forward spectrometer: Technical Design Report*. Technical design report. ALICE. CERN, Geneva, 1999. `https://cds.cern.ch/record/401974`.

[8] **ALICE** Collaboration, "Technical Design Report for the Muon Forward Tracker", tech. rep., 2015. `https://cds.cern.ch/record/1981898`.

[9] M. Mager, "ALPIDE, the Monolithic Active Pixel Sensor for the ALICE ITS upgrade", *Nucl. Instrum. Methods Phys. Res. A* **824** (2016) . `https://www.sciencedirect.com/science/article/pii/S0168900215011122`. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[10] J. H. Friedman, "Greedy function approximation: a gradient boosting machine", *Annals of statistics* (2001) .

[11] T. Sjöstrand *et al.*, "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (Jun, 2015) . `https://doi.org/10.1016%2Fj.cpc.2015.01.024`.

[12] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4–a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) .

[13] K. Albertsson *et al.*, "Machine learning in high energy physics community white paper", *J. Phys. Conf. Ser.* **1085** no. 2, (Sep, 2018) . `https://dx.doi.org/10.1088/1742-6596/1085/2/022008`.

[14] M. Rossi, "Deep Learning Applications to Particle Physics: from Monte Carlo simulation acceleration to ProtoDUNE reconstruction", 2023.

[15] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks", *AIChE Journal* **37** no. 2, (1991) . `https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.690370209`.

[16] **ALICE** Collaboration, "Multiplicity dependence of Υ production at forward rapidity in pp collisions at $\sqrt{s}$ = 13 TeV", 2022.

[17] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library", 2019.

[18] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[19] T. Akiba *et al.*, "Optuna: A next-generation hyperparameter optimization framework", 2019.

[20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., pp. 4765–4774. Curran Associates, Inc., 2017. `http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf`.