

Optimizing K_S^0 reconstruction at different collision energies using Machine Learning algorithms

Facility for Antiproton and Ion Research in Europe and
GSI Helmholtzzentrum für Schwerionenforschung

JULIAN NOWAK



Internship and Training Project Report
Darmstadt, Germany
July-September 2021

Copyrights(c)

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the GSI GmbH and FAIR GmbH.

Author

Julian Nowak
Warsaw University of Technology, Faculty of Physics
00-662 Warsaw, ul. Koszykowa 75, Warsaw - Poland
Email: julian.nowak.stud@pw.edu.pl

Supervisor

Dr. Ilya Selyuzhenkov
CBM-PWG
GSI Helmholtzzentrum für Schwerionenforschung GmbH,
Planckstr. 1, 64291 Darmstadt, Germany
Tel: +49 6159 71 1799
Email: ilya.selyuzhenkov@gmail.com

Program Coordinator

Dr. Pradeep Ghosh
GSI Helmholtzzentrum für Schwerionenforschung GmbH &
Facility for Antiproton and Ion Research in Europe GmbH
Tel: +49 6159 71 3257, Fax: +49 6159 71 3916
Email: Pradeep.Ghosh@fair-center.eu

GET Involved 2021

Publisher: GSI Helmholtzzentrum für Schwerionenforschung GmbH,
Planckstr. 1, 64291 Darmstadt, Germany
Published: September 2021

Abstract

Future Compressed Baryonic Matter (CBM) experiment at FAIR will study the Quantum Chromodynamics (QCD) matter at high density. Short-lived particles which will be produced during the collisions, like K_S^0 , can be reconstructed using i.e. Silicon Tracking System and KFParticle Finder package. To increase the efficiency, and minimize the amount of falsely reconstructed particles, the machine learning algorithms can be used.

In this work, the optimization of K_S^0 reconstruction using XGBoost algorithm was carried out. In the first two chapters, reconstruction procedure and machine learning techniques were presented. In the third chapter, data preparation is described. In the last chapter, the results for $p_{\text{beam}} = 12\text{A}$ GeV/c and $p_{\text{beam}} = 3.3\text{A}$ GeV/c are presented. Also, the influence of magnetic field on the efficiency of the reconstruction was investigated.

Declaration

I hereby declare that the project entitled "**Optimizing K_S^0 reconstruction at different collision energies using Machine Learning algorithms**" is my own work and that I have correctly acknowledged the work of others.

Acknowledgements

I would like to thank for my stay at GSI during the Internship this year, which allowed me to dive into the world of particle physics. It would not be possible without help from Dr. Hanna Zbroszczyk from the Faculty of Physics of Warsaw University of Technology, and Dr. Pradeep Ghosh.

Also, I would like to thank my supervisor, Dr. Ilya Selyuzhenkov, and i.a. Oleksii Lubynets, Shahid Khan, and Olha Lavoryk for much support and help during my training and work on K_S^θ reconstruction.

I hope that I might be able to continue working at CBM as its many challenges seem very interesting and can give an opportunity to learn more for years ahead.

Contents

1	Introduction	12
1.1	CBM experiment	12
1.1.1	CBM experiment	12
1.1.2	CBM setup	13
1.2	Physical motivation	14
1.2.1	Strange particles reconstruction	14
1.2.2	K_S^0 reconstruction	15
2	Short-lived particles reconstruction	16
2.1	KFParticleFinder	16
2.2	PFSimple	16
2.3	Optimization of selection criteria	17
3	Machine Learning	19
3.1	Supervised machine learning	19
3.2	XGBoost	19
3.2.1	Decision trees	19
3.2.2	Comparison of ML with manual selection criteria	20
4	Data preparation	21
4.1	Data enriching	21
4.1.1	Underrepresentation	21
4.1.2	Dependence of collision simulation model	22
4.2	Data cleaning	22
4.2.1	Invariant mass	23
4.2.2	Distances and x , y , z coordinates	23
4.2.3	Momentums	24
4.2.4	Chi square	24
4.2.5	Pseudorapidity	24
4.3	Variables selection	24
4.3.1	Correlation matrix	25

4.3.2	Correlation with invariant mass	25
4.3.3	Selected variables	26
5	Model training	27
5.1	Results for $p_{\text{beam}} = 12A$ GeV/c	27
5.1.1	Cross validation	27
Receiver Operating Characteristic	27	
Probability plot	28	
Invariant mass distribution	29	
5.1.2	Comparison with default KFPF Cuts	30
Better background reduction	30	
Better efficiency	30	
5.1.3	Investigation of potential bias	30
5.2	Results for $p_{\text{beam}} = 3.3A$ GeV/c	34
5.2.1	Comparison with KFPF cuts	34
5.2.2	Influence of the magnetic field scaling	34
6	Summary and outlook	37
7	Bibliography or References	38

List of Figures

1.1	Phase diagram of the QCD matter [2]	12
1.2	CBM setup [3]	13
1.3	Particle population vs density (mixed phase highlighted in yellow) [5]	14
2.1	Decay scheme and topological variables	17
2.2	Example of distributions to be checked with the manual selection criteria optimization	18
3.1	Example of a decision tree using manual selection criteria . . .	20
4.1	Imbalance of classes	21
4.2	Data enriching	22
4.3	Correlation matrix	25
4.4	Correlation of all variables with mass	26
5.1	Receiver Operating Characteristic	28
5.2	Probability plot	29
5.3	Invariant mass distribution	29
5.4	Comparison with KFPF cuts	31
5.5	Invariant mass distribution for probability > 0.86	32
5.6	True-false positives investigation	33
5.7	Comparison with KFPF cuts for 3.3A GeV/c	35
5.8	Invariant mass distribution for different MF scaling	36

List of Tables

2.1	Advantages and disadvantages of manual selection criteria optimization	18
3.1	Advantages and disadvantages of ML selection criteria optimization	20
5.1	Comparison of MF strength vs. efficiency	34

Chapter 1

Introduction

1.1 CBM experiment

1.1.1 CBM experiment

The CBM experiment will be held at the Facility for Antiproton and Ion Research (FAIR) in Darmstadt. Its goal is the exploration of the QCD phase diagram in the region of high net baryon densities using high-energy nucleus-nucleus collisions. It will allow us to i.a. study the equation-of-state of nuclear matter at neutron star core densities. The measurements will be performed at reaction rates up to 10 MHz. It requires highly efficient particle identification and reconstruction framework.[1]

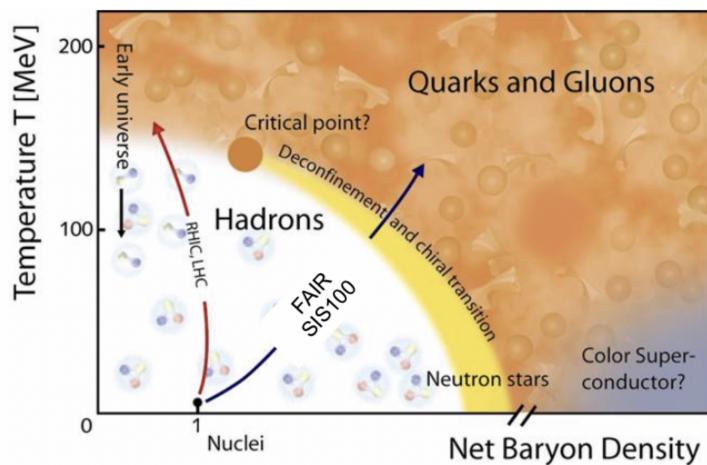


Figure 1.1: Phase diagram of the QCD matter [2]

1.1.2 CBM setup

In order to identify the particles coming from the collisions, the setup of 8 detectors is under development in GSI Darmstadt[2].

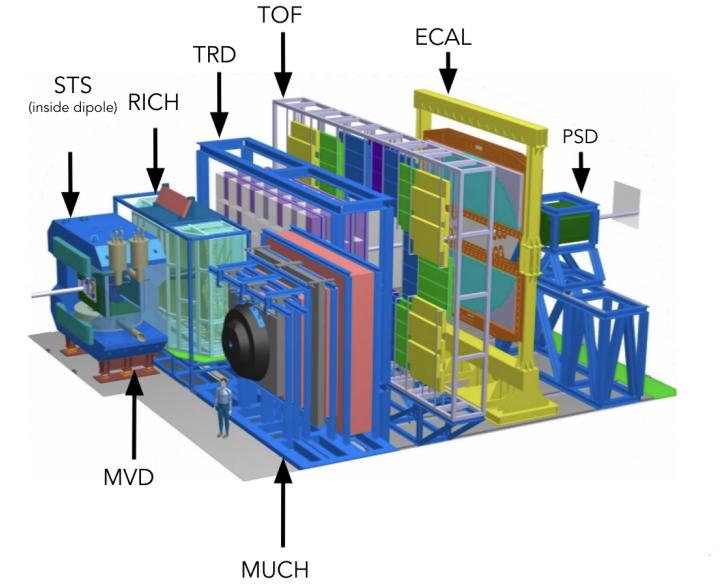


Figure 1.2: CBM setup [3]

The CBM detectors setup will consist of:

- STS - Silicon Tracking System
- MVD - Micro Vertex Detector
- RICH - Ring Imaging Cherenkov Detector, replaceable with:
- MUGH - Muon Chamber System
- TRD - Transition Radiation Detector
- TOF - Time of Flight Detecotor
- ECAL - Electromagnetic Caloromiter
- PSD - Projectile Spectator Detector

While each detector has its own role, in this work I will mainly focus on the results coming from two of them: charged particles like protons or pions can

be identified using the Time of Flight Method[4], which bases on the TOF detector; the short-lived particles reconstruction depends mainly on the STS detector.

1.2 Physical motivation

1.2.1 Strange particles reconstruction

In the nuclei collisions in experiments like CBM or HADES, the strange quarks are only produced during collisions. Thus, they provide information about the evolution of nuclear matter. We expect [5] that in the *mixed phase* (when both nucleon and quark degrees of freedom are present), which can be produced at FAIR energies, the yield of strange particles could be comparable with particles composed of u and d quarks. To verify this, we need to reconstruct all the strange particles, also short-lived strange particles like Λ^0 , Ξ^- , K_S^0 which cannot be detected directly.

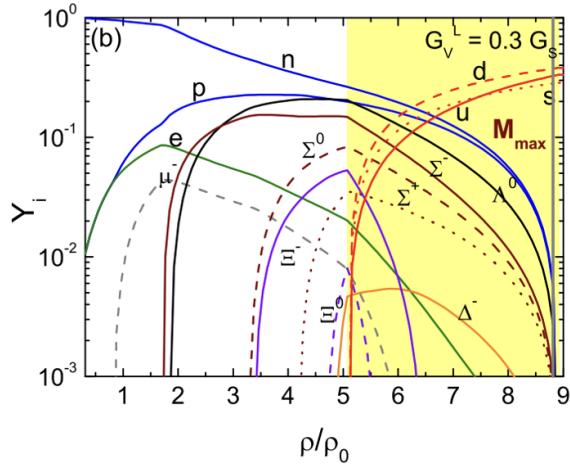
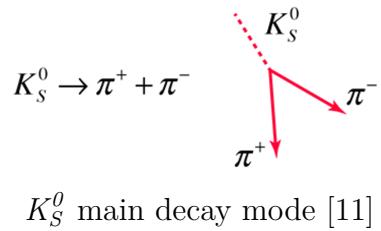


Figure 1.3: Particle population vs density (mixed phase highlighted in yellow) [5]

1.2.2 K_S^0 reconstruction

In this work, I will mainly focus on the reconstruction of one of the short-lived particles, K-short (short-lived Kaon):

- Mass: $497.611 \pm 0.013 MeV/c^2$
- Mean lifetime: $8.958 \cdot 10^{-11} s$
- Charge = 0
- Meson, composed of two quarks: $d \bar{s}$ or $s \bar{d}$
- Strange particle
- in the main decay mode (69.2% of cases) decays into π^+ and π^- [12]



The main decay mode of K_S^0 is more symmetric than e.g., Λ^0 main decay mode (decay into p and π^-); its products are also less massive than the ones coming from Λ^0 decay. Therefore, the investigation of this decay can tell us much about the tracking performance in CBM. Also, we can explore the influence of the magnetic field strength on the efficiency of the reconstruction of short-lived particles.

Chapter 2

Short-lived particles reconstruction

2.1 KParticleFinder

Short-lived particles like Λ^0 , and K_S^0 (mean lifetime respectively: $2.631 \cdot 10^{-10}$ s, and $8.958 \cdot 10^{-11}$ s) have neutral charge, therefore they cannot be detected directly. However, we can reconstruct these particles by investigating their *daughter particles*, particles coming from their decays. Using i.e. the Silicon Tracking System we can reconstruct the tracks of pions coming from K_S^0 decay. In order to do so, the *KParticleFinder* was developed [13]. It is an on-line optimized reconstruction package based on Kalman Filter mathematics. It finds pairs of positive and negative pions (in our case) which could be a result of K-short decay.

2.2 PFSimple

For offline selection optimization and analysis, PFSimple package was created [14]. Using it, one can optimize selection criteria to differentiate between:

- signal - pion pairs created in K-short decay
- background - pions pairs returned by KParticleFinder which are not the result of K-short decay

Also, PFSimple returns these attributes of each reconstructed particle:

- $L/\Delta L$ - distance between primary and secondary vertex divided over its error

- DCA - distance of closest approach between pion tracks
- χ^2 geo - dimensionless distance of closest approach between pion tracks
- χ^2 topo - dimensionless distance of closest approach between extrapolated kaon trajectory and primary vertex
- χ^2 prim - dimensionless distance between extrapolated secondary track and primary vertex
- $\cos(\alpha)$ - cosine of angle between pion and K-short momenta

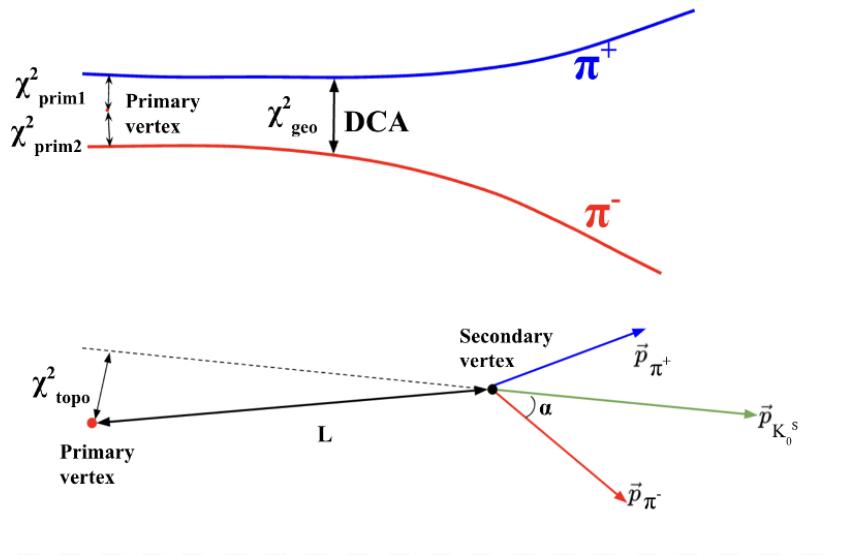


Figure 2.1: Decay scheme and topological variables

2.3 Optimization of selection criteria

As the KFParticle Finder (and PFSimple) return the possible pion pairs, which could be a result of K-short decay, some selection criteria need to be applied to differentiate between the signal and background.

Manual procedure for selection criteria optimization [15] is as follows:

- The goal is to suppress as a lot background and to preserve as much signal as possible
- Plot distribution of signal and background for some variable, and select a point above which all entries are considered as signal or background

- Repeat it for all the variables

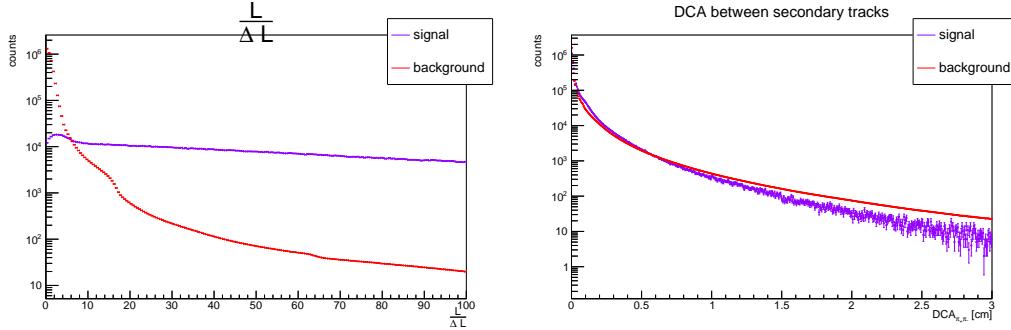


Figure 2.2: Example of distributions to be checked with the manual selection criteria optimization

This method has the following advantages and drawbacks:

Advantages	Disadvantages
Easily interpretable	Linear i.e. one dimension/variable is used once
Computationally inexpensive	Non-automatized
	Collision simulating model (e.g. UrQMD) dependent

Table 2.1: Advantages and disadvantages of manual selection criteria optimization

Another approach is using *Machine Learning* to optimize selection criteria, which is the main topic of this work, and will be described in the next chapters.

Chapter 3

Machine Learning

3.1 Supervised machine learning

Supervised machine learning, is "a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled datasets to train algorithms that classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross validation process." [16]

3.2 XGBoost

XGBoost is a decision-tree-based (supervised) Machine Learning algorithm that uses gradient boosting. [17]. It is highly efficient when used with e.g., tabular data, so it suits the need for optimization selection criteria. [18]

3.2.1 Decision trees

We can also visualize manual selection criteria with a decision tree:

1. For each variable of a reconstructed particle, a conditional statement is being set
2. If all conditions are fulfilled, a K-short candidate is treated as a signal candidate

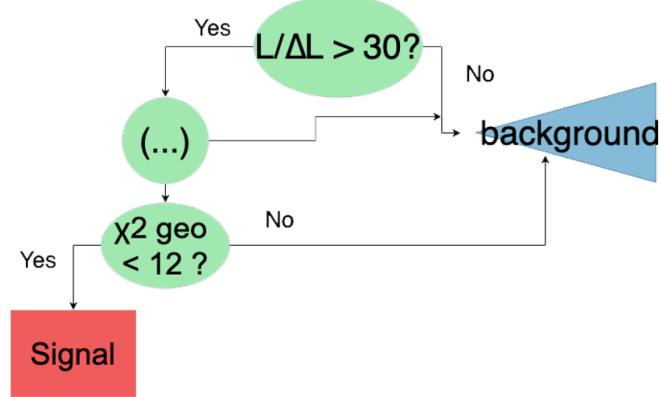


Figure 3.1: Example of a decision tree using manual selection criteria

In XGBoost, the algorithm creates multiple decision trees, combines them, and returns *probability* whether a K-short candidate is a signal or background candidate.

3.2.2 Comparison of ML with manual selection criteria

Advantages	Disadvantages
Non-linear in multi-dimensional space	Not easily interpretable
Automatization of the selection process	Computationally expensive
Partially collision simulating model-independent With <i>probability</i> a value from which we consider a particle as a K-short particle can be chosen; it can aim for better reconstruction efficiency or a better background reduction	

Table 3.1: Advantages and disadvantages of ML selection criteria optimization

Chapter 4

Data preparation

4.1 Data enriching

There are two problems to tackle before providing datasets to our ML model: underrepresentation and dependence of collision simulation model.

4.1.1 Underrepresentation

As in the simulated data, less than 0.1% of K-short candidates are signal candidates, the ratio in the training dataset has to be changed so that the model learns how to identify signal candidates well. The number of background entries is rescaled by this formula:

$$n_{\text{events}(\text{background})} = 5 \cdot n_{\text{events}(\text{signal})} \quad (4.1)$$

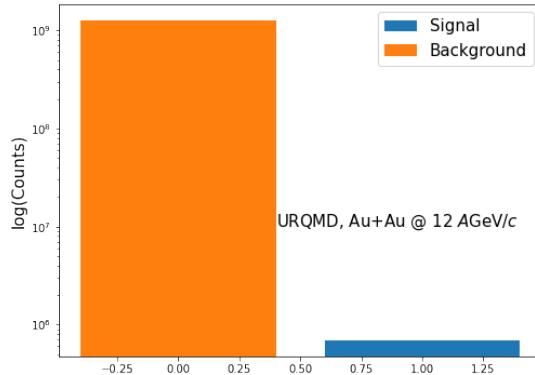


Figure 4.1: Imbalance of classes

4.1.2 Dependence of collision simulation model

To make the ML model partially independent of the collision simulation model, select:

- primary K_S candidates only in 5σ region: $0.43485 - 0.56135$ (GeV/c^2) from DCM-QGSM-SMM model[6][7], (simulation data)
- background outside 5σ region from UrQMD model[8], which mimics experimental data (will be replaced by real experimental data while CBM experiment starts)

Also, the UrQMD model is being used as a test dataset (as the model does not "know" the signal candidates from this MC model).

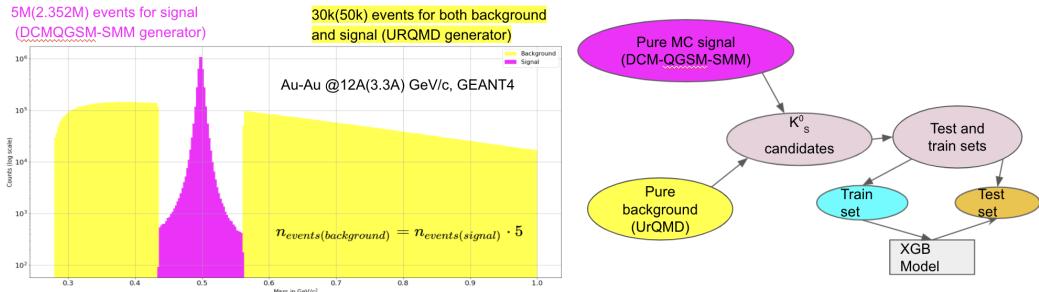


Figure 4.2: Data enriching

In total, the following datasets will be used:

- 5M (2.352M) events for signal generated in DCM-QGSM-SMM
- 30k(50k) events for both background and test dataset generated in UrQMD

Au-Au @12A(3.3A) GeV/c passed through CBM setup in GEANT4.[9][10]

4.2 Data cleaning

To reject the numeric values of parameters which do not have physical sense, but are present in the data set, some selection criteria are applied before the beginning of the model training. Similarly, we some values which might be possible but are rare enough are rejected, to reduce the amount of data.

4.2.1 Invariant mass

As the K-short particle decays into two pions, its invariant mass cannot be smaller than the mass of the two pions, so:

$$m_{\text{inv}} > 0.279 \text{ GeV}/c^2$$

Also, to reduce the amount of data:

$$m_{\text{inv}} < 1 \text{ GeV}/c^2$$

4.2.2 Distances and x, y, z coordinates

Distance between the primary vertex (the point where the collision of the nuclei happens), and the secondary vertex (the extrapolated point where the two daughter particles should have crossed each other) - l and the distance of closest approach between the two pions - DCA - should not be smaller than zero:

$$DCA, l, \frac{l}{\Delta l} > 0$$

Also, due to the sizes of the tracking system (the largest station has an area of 1m^2):

$$DCA < 100 \text{ cm}$$

For the same reason:

$$|x|, |y| < 50 \text{ cm}$$

As the particle has to hit 3 stations of the tracking system, and the last two are placed above 80 cm:

$$l < 80\text{cm}$$

For the same reason, and because of the fixed target geometry of the detector:

$$-1 \text{ cm} < z < 80 \text{ cm}$$

To reduce the data, we set:

$$\frac{l}{\Delta l} < 15000$$

However, in the KFParticle package l is assumed to be signed by design, and one can notice that actually, some data entries have a negative value of distance, for both signal and background. As the *quality cuts* should be rather conservative, the following ranges are set:

$$\begin{aligned} l &> -5 \text{ (cm)} \\ \frac{l}{\Delta l} &> -25 \end{aligned}$$

4.2.3 Momentums

The fixed target geometry of the detector requires that:

$$p_Z > 0 \text{ GeV/c}$$

To reduce the data:

$$p < 20 \text{ GeV/c}; p_T < 3 \text{ GeV/c}$$

4.2.4 Chi square

Since χ^2 is a squared distance, all the values must be larger than zero:

$$\chi^2 > 0$$

To reduce the data, following the maximal values are selected:

- χ^2 first and second $< 3 \cdot 10^7$
- $\chi_{geo}^2 < 10000$
- $\chi_{topo}^2 < 100000$

4.2.5 Pseudorapidity

As pseudorapidity depends on polar angle as $\eta = -\ln \tan(\frac{\theta}{2})$, and the Silicon Tracking System (STS) covers the polar angles between 2.5° and 25° , for which the pseudorapidity values would equal:

$$1.5 < \eta < 3.82$$

However, due to the magnetic field, the pseudorapidity is constrained to following values:

$$1.0 < \eta < 6.5$$

with which we loose 0.06% of data for signal (instead of 5.66%) and 0.08% of data for background (instead of 6.75%)

4.3 Variables selection

The last step before training our model is the selection of variables which will be used in the training of the discriminator. To make the model simpler, there is no need to use two variables if they are strongly correlated with each other. Also, to avoid signal/background classification directly by invariant mass, variables strongly correlated with the invariant mass of background should be omitted.

4.3.1 Correlation matrix

The Pearson correlation efficient is being calculated following this formula:

$$\rho = \frac{\text{COV}(X, Y)}{\sigma_X \times \sigma_Y} \quad (4.2)$$

where $\text{COV}(X, Y) = E[(X - E[X])(Y - E[Y])]$; $\sigma_X = \sqrt{E[(X - E[X])^2]}$ of each variable and plot it in a matrix:

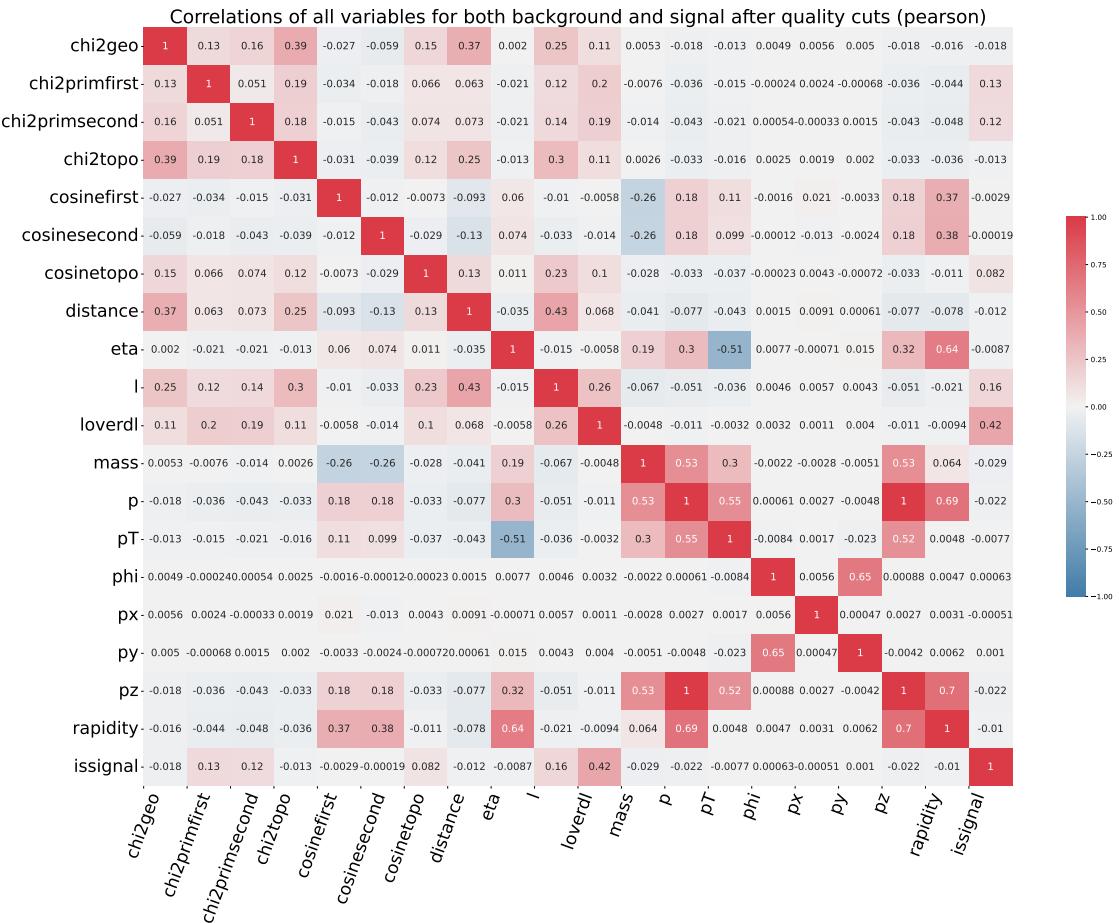


Figure 4.3: Correlation matrix

4.3.2 Correlation with invariant mass

Later, the correlation (Pearson coefficient) of each variable with the invariant mass of (separately) signal and background is being checked

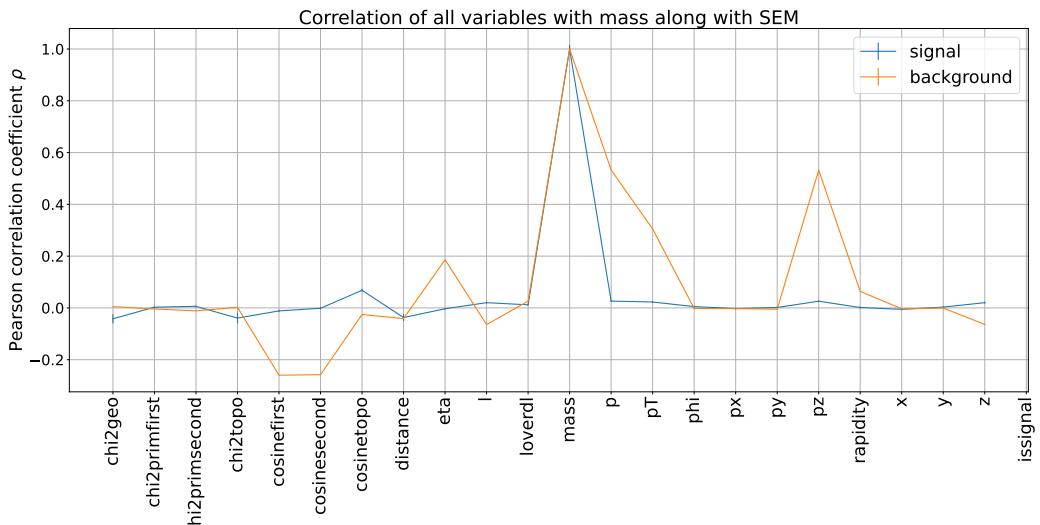


Figure 4.4: Correlation of all variables with mass

Also, a study of non-linear correlations has been performed. [19]

4.3.3 Selected variables

Based on that, the following 6 variables are selected for the training: $L/\Delta L$, DCA ; χ^2_{geo} , χ^2_{topo} , $\chi^2_{primfirst}$, $\chi^2_{primsecond}$

Chapter 5

Model training

5.1 Results for $p_{\text{beam}} = 12\text{A}$ GeV/c

Having prepared the data, the training of the ML algorithm may be initiated. To choose the hyper-parameters of the XGBoost model, the Bayesian Optimisation package was used.[20]

5.1.1 Cross validation

To check if the model works as well on new data, as on the training dataset, Receiver Operating Characteristic and probability plot are prepared.

Receiver Operating Characteristic

Receiver Operating Characteristic illustrates the diagnostic ability of a binary classifier. Threshold on the ROC (Receiver Operating Characteristic) curve which maximizes Approximate Median Significance

$$\text{AMS} = \sqrt{2}[(tpr + fpr) \log(1 + tpr/fpr) - tpr] \quad (5.1)$$

(where $t(f)pr$ is true (false) positive rate) on the test sample is the best threshold.

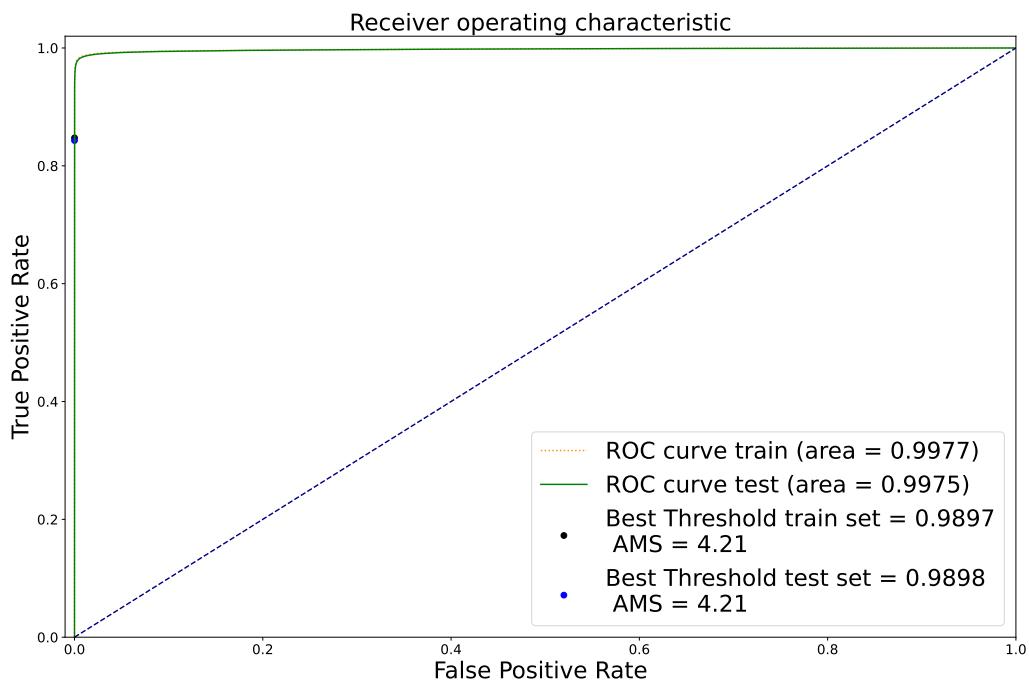


Figure 5.1: Receiver Operating Characteristic

We see that the optimal point on ROC is similar for both train and test datasets.

Probability plot

The graph of signal/background share in both train and test datasets is also plotted.

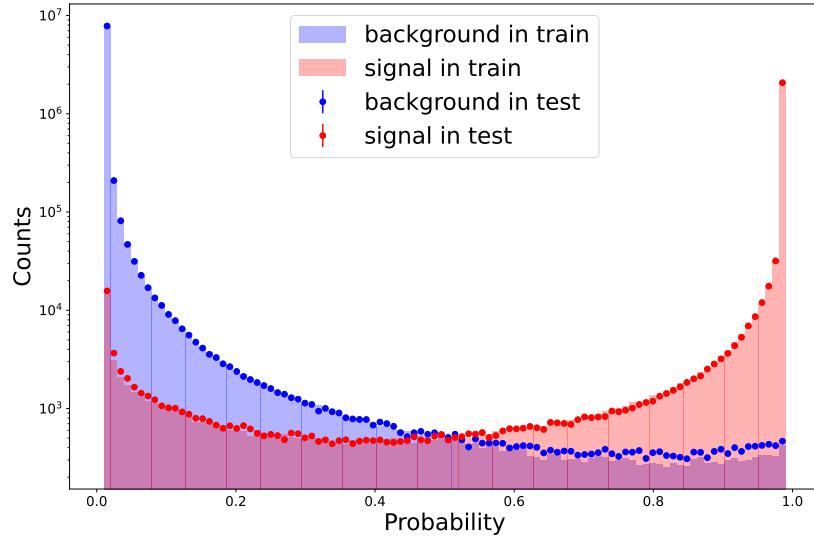


Figure 5.2: Probability plot

Once again, the results are similar for the two datasets. Hence, we can say that our model is not overtrained (works well on new data).

Invariant mass distribution

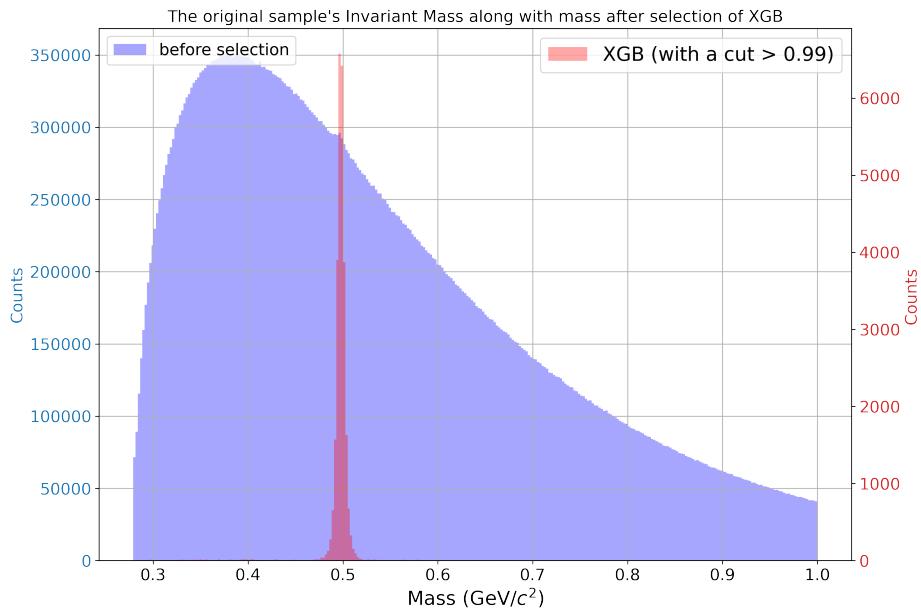


Figure 5.3: Invariant mass distribution

5.1.2 Comparison with default KFPF Cuts

KFParticleFinder has default selection criteria, which can be compared with ML selection:

- $L/\Delta L > 5$
- $DCA < 1 \text{ cm}$
- $\chi_{geo}^2 < 3$
- $\chi_{prim}^2 > 18.4$
- $\cos(\alpha) > 0$

Depending on the cut value, better background reduction or better efficiency can be obtained.

Better background reduction

With a cut value set to 0.99, we get $50\times$ less background (Fig. 5.4):

- **Reconstructed K_S^0 / reconstructible $K_S^0 = 80.67\%$ vs. 76.93% with default KFPF cuts**
- **false / true positive rate = 0.04** vs. 2.01 with default KFPF cuts

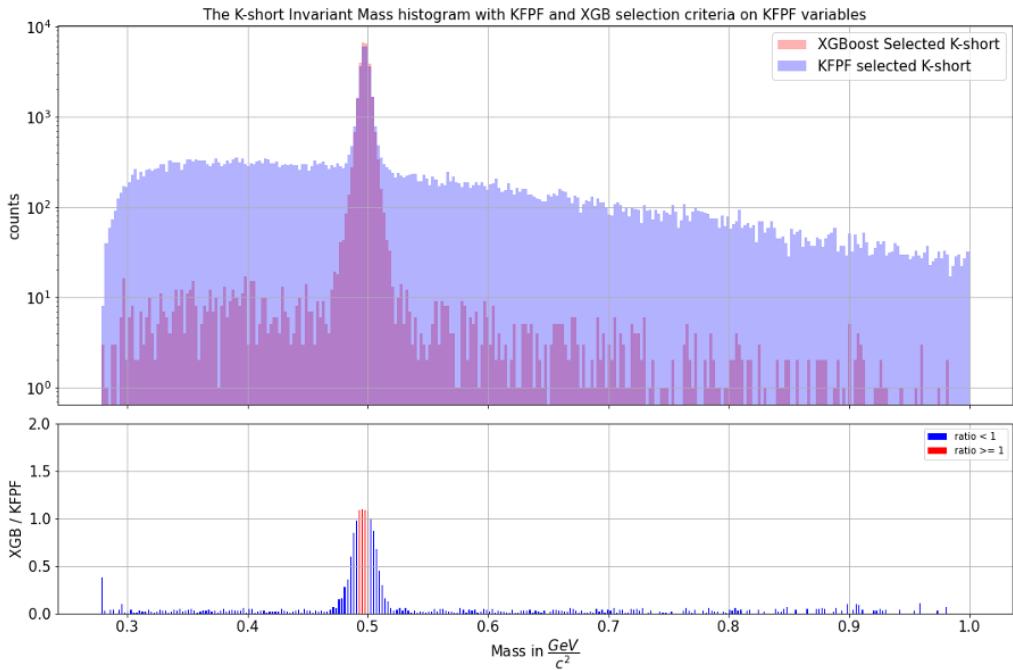
Better efficiency

With a cut value set to 0.86, we get 20% better efficiency (Fig. 5.5):

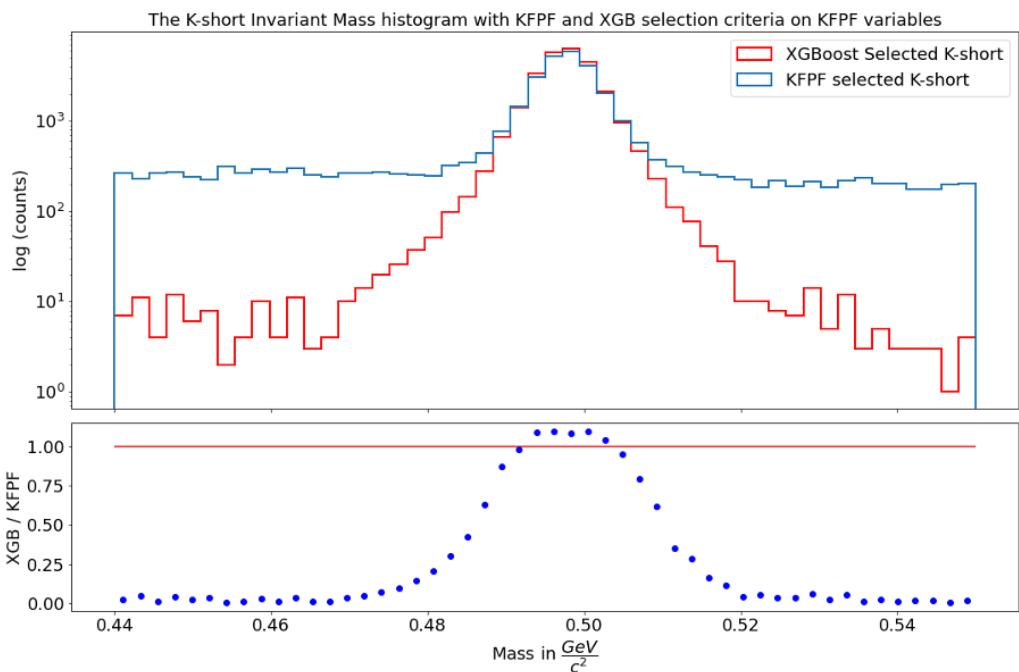
- **Reconstructed K_S^0 / reconstructible $K_S^0 = 95.91\%$ vs. 76.93% with default KFPF cuts**
- **false / true positive rate = 1.27** vs. 2.01 with default KFPF cuts

5.1.3 Investigation of potential bias

To check if the XGB selection criteria cut tails of the distribution, or are biased in some regions, the invariant mass distribution of true and false positives is plotted (Figure. 5.6). We see that our model does not seem to be biased towards any invariant mass region. Also, an investigation of influence of XGBoost on each variable has been performed.[21]

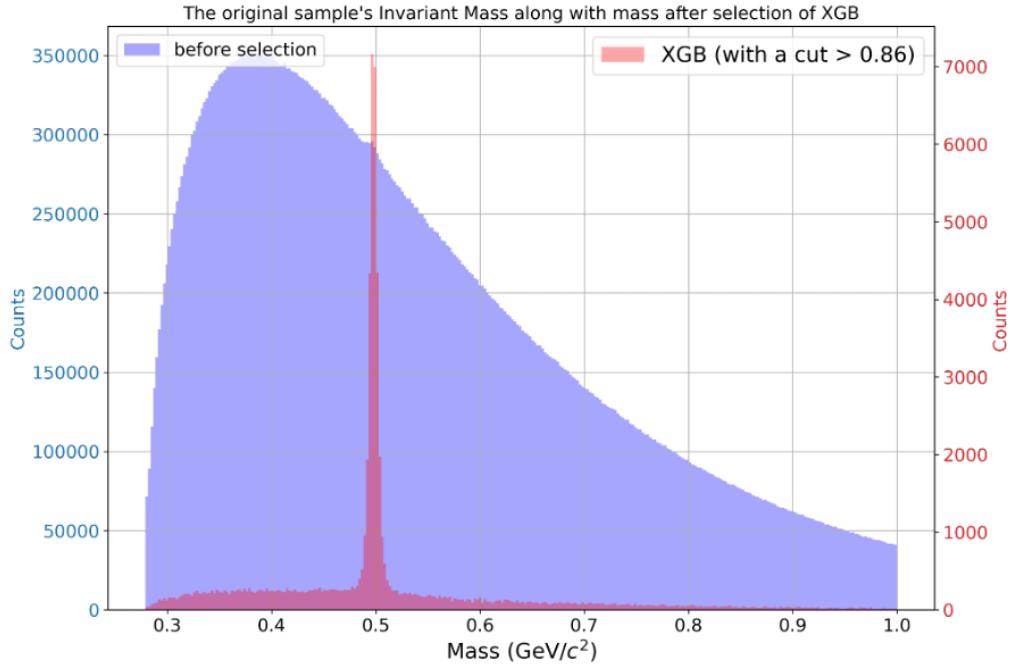


(a) Invariant mass distribution (y log scale)

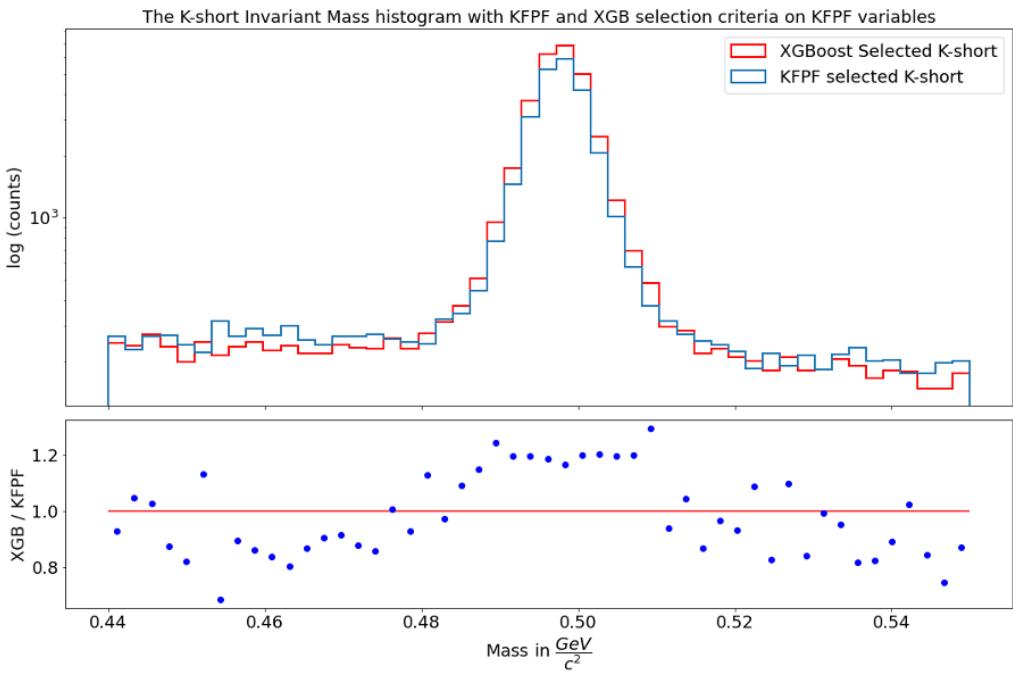


(b) Invariant mass distribution (close-up)

Figure 5.4: Comparison with KFPF cuts

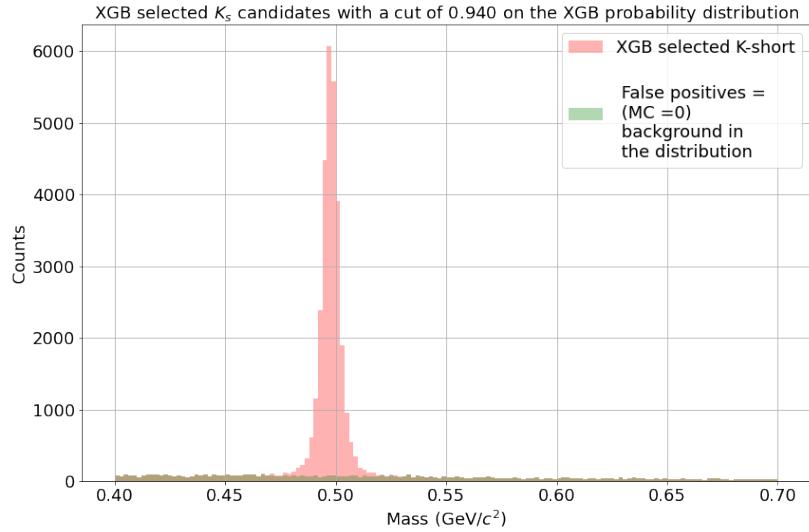


(a) Invariant mass distribution (y log scale)

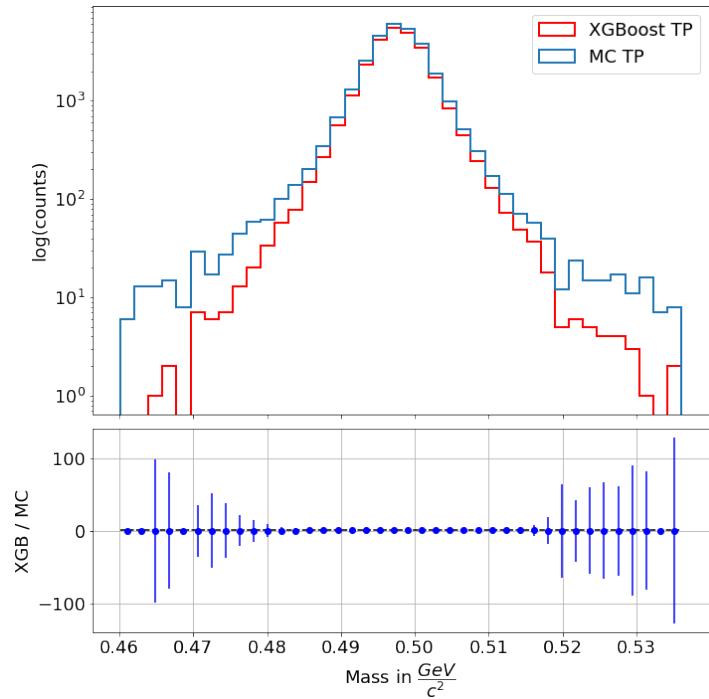


(b) Invariant mass distribution (close-up) - comparison with KFPF cuts

Figure 5.5: Invariant mass distribution for probability > 0.86



(a) Selected K_S^θ and false positives distribution



(b) Reconstructed and MC true positives (y log scale)

Figure 5.6: True-false positives investigation

5.2 Results for $p_{\text{beam}} = 3.3\text{A GeV/c}$

5.2.1 Comparison with KFPF cuts

The same code can be used to obtain similar results for another CBM energy level. Comparing to default KFPF cuts, with probability cut 0.9635 (Fig. 5.7):

- Reconstructed K_S^0 / reconstructible $K_S^0 = 93.45\%$ vs. 78.94% with default KFPF cuts
- false / true positive rate = 0.19 vs. 1.38 with default KFPF cuts

Due to the smaller statistics for this energy level, the ML model training should be redone for a bigger dataset.

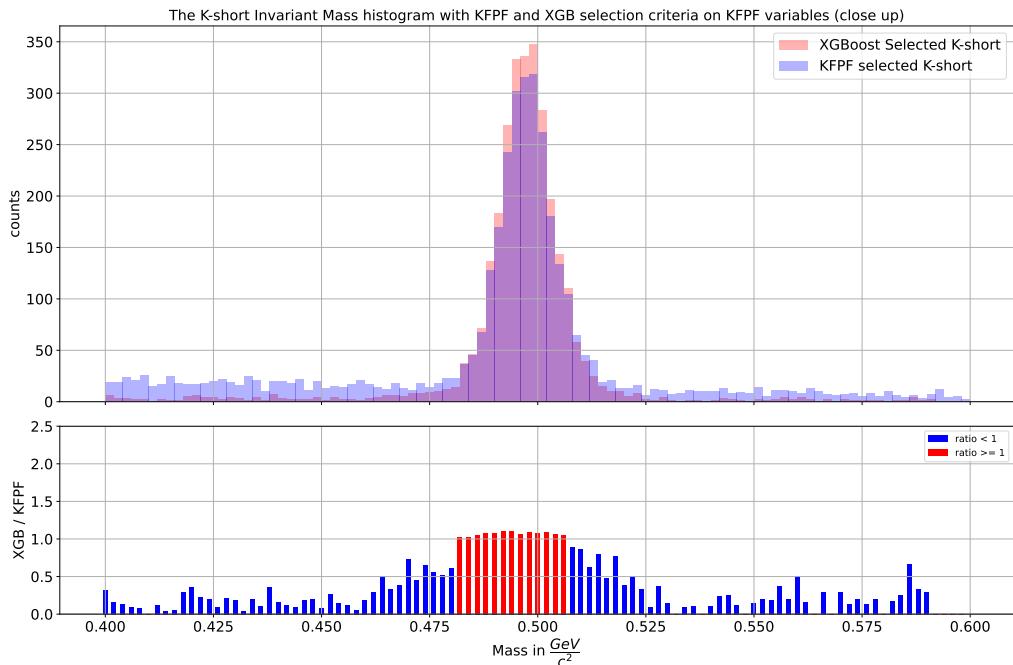
5.2.2 Influence of the magnetic field scaling

Obtained ML model can be adapted for the investigation of the influence of the magnetic field scaling. We compare: 100% MF strength (for $p_{\text{beam}} = 12\text{A GeV/c}$), 56% and 27.5% MF strength (for $p_{\text{beam}} = 3.3\text{A GeV/c}$, both with only DCM generated data for 0.4M events for training and 0.1M for validation). We see (Fig. 5.8) that the stronger the magnetic field is, the broader K_S^0 invariant mass distribution peak is. Also, we observe much more signal entries for $p_{\text{beam}} = 12\text{A GeV/c}$ (for the same number of events).

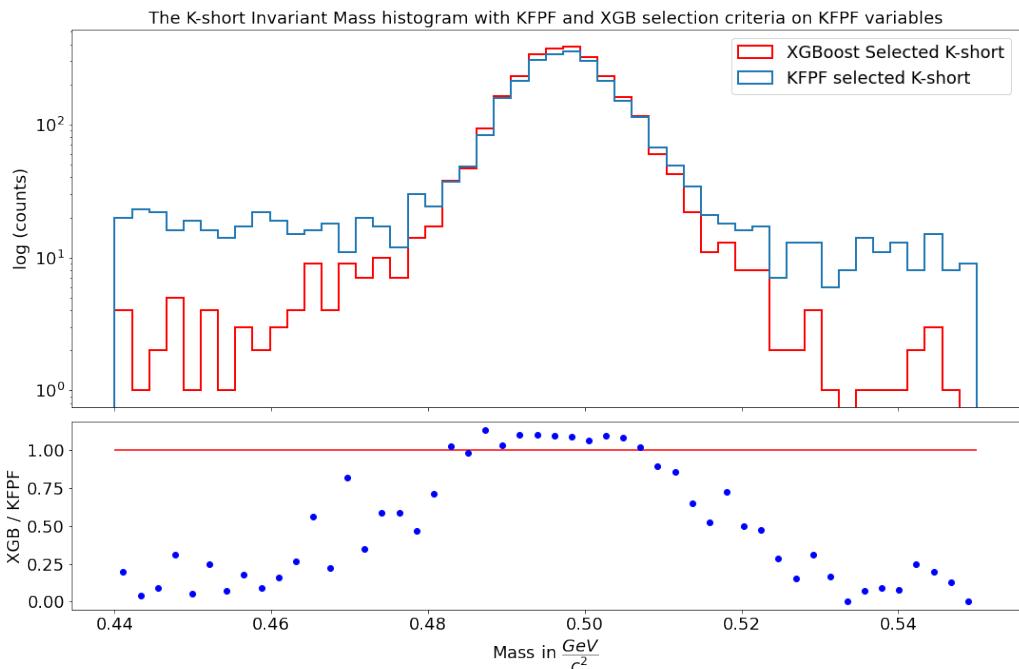
For $p_{\text{beam}} = 3.3\text{A GeV/c}$ the number of signal entries is almost the same; we select the probability cut so that the false/true positive ratio is the same for both % of MF and compare the efficiency of the reconstruction (Tab. 5.1). We see that weaker MF does not necessarily worsen efficiency (only 2% difference). However, this comparison should also be redone for a bigger dataset.

ratio	12 A GeV/c MF=100%	3.3 A GeV/c MF=56%	3.3 A GeV/c MF=27.5%
reconstructed/ reconstructible	89.98%	90.36%	88.49%
false / true positive	0.2	0.2	0.2

Table 5.1: Comparison of MF strength vs. efficiency



(a) Invariant mass distribution (y log scale)



(b) Invariant mass distribution (close-up)

Figure 5.7: Comparison with KFPF cuts for 3.3A GeV/c

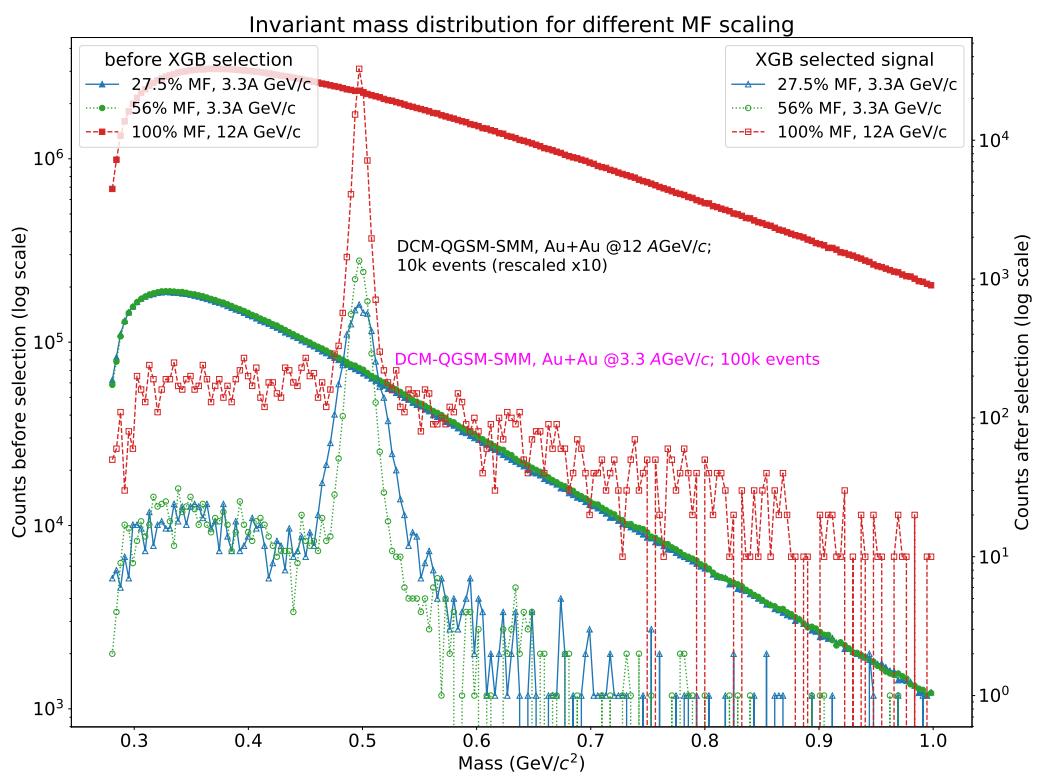


Figure 5.8: Invariant mass distribution for different MF scaling

Chapter 6

Summary and outlook

The optimization of K_S^0 reconstruction at different collision energies ($p_{\text{beam}} = 12\text{A GeV}/c$ and $3.3\text{A GeV}/c$) was performed. After preparing simulated data (enriching, cleaning, and variables selection) and loading them into XGBoost framework, we obtain a well working ML model which allows us to differentiate between the *signal* - pion pairs created in K-short decay, and *background* - pions pairs returned by KFParticle Finder, which are not the result of a K_S^0 decay. We observe that a higher collision energy results in a bigger K_S^0 production. Also, the investigation of influence of the magnetic field was performed - the efficiency of the reconstruction does not seem to be lower with a weaker MF; this investigation should be redone with a bigger dataset though.

Created framework can be applied to perform a multi-differential analysis of the selection optimization in transverse momentum, rapidity, and centrality at different collision energies. Also, the discriminator should be prepared to be imported to AnalysisTree in (Cbm)Root.

Chapter 7

Bibliography or References

Bibliography

- [1] <https://fair-center.eu/for-users/experiments/cbm.html>
- [2] "Compressed Baryonic Matter Experiment at FAIR, Progress Report 2020" Senger, Peter et. al (CBM Collaboration) doi:10.15120/GSI-2021-00421.
- [3] https://indico.gsi.de/event/13089/contributions/56358/attachments/37175/49720/38th_CBM_CollaborationMeeting_KarabowiczMalgorzata_MLforXi-Reconstruction.pdf
- [4] M. C. S. Williams, "Particle identification using time of flight," J. Phys. G **39** (2012), 123001 doi:10.1088/0954-3899/39/12/123001
- [5] M. Orsaria, H. Rodrigues, F. Weber and G. A. Contrera, "Quark deconfinement in high-mass neutron stars," Phys. Rev. C **89** (2014) no.1, 015806 doi:10.1103/PhysRevC.89.015806 [arXiv:1308.1657 [nucl-th]].
- [6] M. Baznat, A. Botvina, G. Musulmanbekov, V. Toneev and V. Zhezher, "Monte-Carlo Generator of Heavy Ion Collisions DCM-SMM," Phys. Part. Nucl. Lett. **17** (2020) no.3, 303-324 doi:10.1134/S1547477120030024 [arXiv:1912.09277 [nucl-th]].
- [7] A. S. Botvina, I. N. Mishustin, M. Begemann-Blaich, J. Hubelle, G. Imme, I. Iori, P. Kreutz, G. J. Kunde, W. D. Kunze and V. Lindenstruth, *et al.* "Multifragmentation of spectators in relativistic heavy ion reactions," Nucl. Phys. A **584** (1995), 737-756 doi:10.1016/0375-9474(94)00621-S
- [8] <https://urqmd.org/>

- [9] S. Agostinelli *et al.* [GEANT4], “GEANT4—a simulation toolkit,” Nucl. Instrum. Meth. A **506** (2003), 250–303 doi:10.1016/S0168-9002(03)01368-8
- [10] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. A. Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie and R. Chytracek, *et al.* “Geant4 developments and applications,” IEEE Trans. Nucl. Sci. **53** (2006), 270 doi:10.1109/TNS.2006.869826
- [11] <http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/kaon.html>
- [12] P. A. Zyla *et al.* [Particle Data Group], “Review of Particle Physics,” PTEP **2020** (2020) no.8, 083C01 doi:10.1093/ptep/ptaa104
- [13] https://indico.cern.ch/event/773049/contributions/3476146/attachments/1939866/3218991/Kisel_CBM_CHEP-2019.pdf
- [14] https://indico.gsi.de/event/10062/contributions/48045/attachments/33320/43053/PFSimple4Lambda_HIGHLIGHTS_36CM.pdf
- [15] https://indico.gsi.de/event/10126/contributions/362/attachments/229/363/PFsimple_cuts_MC-optimization.pdf
- [16] <https://www.ibm.com/cloud/learn/supervised-learning>
- [17] <https://xgboost.ai>
- [18] S. Khan, V. Klochkov, O. Lavyryk, O. Lubynets, A. I. Khan, A. Dubla and I. Selyuzhenkov, “Machine Learning Application for Λ Hyperon Reconstruction in CBM at FAIR,” [arXiv:2109.02435 [physics.ins-det]].
- [19] https://github.com/julnow/JupyterNotebooks/blob/kaon/img/correlations/bckgr_mass.pdf.pdf
- [20] <https://github.com/fmfn/BayesianOptimization/>
- [21] https://github.com/julnow/JupyterNotebooks/blob/kaon/img/xgb_12agev/chi2geo/params.pdf