

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ  
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
"НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ  
"ВЫСШАЯ ШКОЛА ЭКОНОМИКИ"  
ФАКУЛЬТЕТ КОМПЬЮТЕРНЫХ НАУК**

Шипилов Фома Александрович

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ  
ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕКОНСТРУКЦИИ  
ФОКУСИРУЮЩЕГО АЭРОГЕЛЕВОГО ДЕТЕКТОРА  
MACHINE LEARNING APPLICATION FOR FOCUSING AEROGEL  
DETECTORS IN HIGH ENERGY PHYSICS**

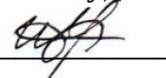
по направлению подготовки 01.04.02 Прикладная математика и информатика  
образовательная программа «Математика машинного обучения»

Научный руководитель  
к.ф.-м.н., Ведущий научный сотрудник



Ф.Д. Ратников

Студент



Ф.А. Шипилов

Москва 2025

**FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION FOR  
HIGHER PROFESSIONAL EDUCATION NATIONAL RESEARCH  
UNIVERSITY  
«HIGHER SCHOOL OF ECONOMICS»  
*Faculty of Computer Science***

---

Foma Shipilov

---

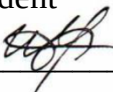
**Применение машинного обучения для реконструкции фокусирующего  
аэрогелевого детектора**

---


**Machine Learning Application for Focusing Aerogel Detectors in High Energy  
Physics**

Qualification paper – Master of Science Dissertation  
Field of study 01.04.02 «Applied Mathematics and Informatics»  
Program: Math of Machine Learning

Student

  
Foma Shipilov

Supervisor

  
Fedor Ratnikov

Moscow, 2025

# **Machine Learning Application for Focusing Aerogel Detectors in High Energy Physics**

Foma Shipilov

*Submitted to the Skolkovo Institute of Science and Technology on March 28, 2025*

## **ABSTRACT**

In the end-cap region of the SPD detector complex, particle identification will be provided by a Focusing Aerogel RICH detector (FARICH). FARICH's primary function is to separate pions and kaons in final open charmonia states (momenta below 5 GeV/c). The optimization of detector parameters, as well as a free-running (triggerless) data acquisition pipeline to be employed in the SPD necessitate a fast and robust method of event reconstruction. In this work, we employ a Convolutional Neural Network (CNN) for FARICH particle identification. The CNN model achieves a superior separation between pions and kaons compared with traditional approaches. Unlike algorithmic methods, an end-to-end CNN model is able to process a full 2-dimensional detector response and skip the intermediate step of computing particle velocity, solving the particle classification task directly.

Keywords: NICA SPD, FARICH, machine learning, particle identification, detector reconstruction

Research advisor:

Name: Fedor Ratnikov

Degree, title: PhD, Lead Scientist

# **Machine Learning Application for Focusing Aerogel Detectors in High Energy Physics**

Foma Shipilov

*Submitted to the Skolkovo Institute of Science and Technology on March 28, 2025*

## **ABSTRACT**

Для улучшения идентификации частиц в торцевом регионе установки SPD будет установлен детектор Focusing Aerogel RICH (ФАРИЧ). Основная задача ФАРИЧ состоит в разделении пионов и каонов из завершающих состояний чармония (импульсы до 5 ГэВ/с). Задачи разработки детектора, а так же бестриггерный режим работы системы сбора данных требуют создания быстрого и надежного метода реконструкции событий. В данной работе авторы применяют для задачи идентификации частиц в ФАРИЧ сверточные нейронные сети (CNN). Сверточная модель демонстрирует более высокое качество разделения пионов и каонов по сравнению с алгоритмическими подходами. В отличие от них, сверточная сеть, обучаемая в режиме полной автономности (от одного конца до другого), получает на вход двумерные данные и пропускает промежуточный этап вычисления скорости частицы, что позволяет напрямую решать задачу классификации частиц.

Keywords: NICA SPD, FARICH, machine learning, particle identification, detector reconstruction

Research advisor:

Name: Fedor Ratnikov

Degree, title: PhD, Lead Scientist

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
<b>2</b>	<b>Literature Review</b>	<b>7</b>
2.1	RICH PID . . . . .	7
2.2	Current Analogous Methods . . . . .	8
2.3	Relation to Current State of the Art . . . . .	8
2.4	NICA SPD Details . . . . .	9
<b>3</b>	<b>Problem Statement</b>	<b>10</b>
3.1	FARICH Reconstruction . . . . .	10
3.2	End-to-End Optimization Proposal . . . . .	11
<b>4</b>	<b>Methods</b>	<b>12</b>
4.1	Refraction Correction . . . . .	12
4.2	Algorithmic Methods . . . . .	14
4.3	Machine Learning Models . . . . .	14
<b>5</b>	<b>Numerical Experiments</b>	<b>17</b>
<b>6</b>	<b>Results and Discussion</b>	<b>19</b>
6.1	Regression Metrics . . . . .	19
6.2	Classification Metrics . . . . .	20
6.3	Discussion . . . . .	21
<b>7</b>	<b>Conclusions</b>	<b>23</b>
<b>8</b>	<b>Author Contribution</b>	<b>24</b>
<b>9</b>	<b>List of publications</b>	<b>25</b>
	<b>Acknowledgements</b>	<b>26</b>
	<b>Bibliography</b>	<b>27</b>
	<b>Appendix A</b>	<b>30</b>
	<b>Appendix B</b>	<b>32</b>

# Chapter 1

## Introduction

The Spin Physics Detector (SPD) is a universal detector proposed by the SPD collaboration at the Nuclotron-based Ion Collider fAcility (NICA) to study the Drell–Yan (DY) processes,  $J/\Psi$  production processes, elastic reactions, spin effects in one and two hadron production processes, polarization effects in heavy ion collisions, and more (Fig. 1.1, left) [29]. The SPD is a medium energy experiment, offering unique possibilities of beam operation and bridging the gap between the low-energy measurements, e.g. ANKE-COSY [22] and high-energy measurements, such as Relativistic Heavy Ion Collider [23]. Several unusual design characteristics arise from the unique goals of the project. A high luminosity up to  $10^{32} \text{ cm}^{-2}\text{s}^{-1}$  and a free-flowing (triggerless) running mode require novel approaches to the data acquisition [1].

The experimental high-energy physics (HEP) objectives require searching for rare signals in background dominated environments. Machine learning techniques can extract high-level representations from the input data and model complex relations while providing excellent scalability and ease of parallelization. Machine learning approaches are state-of-the-art across a diversity of HEP problems [3].

Reliable particle identification (PID) is a crucial component of modern HEP experiments. Originally, a Time-of-Flight (ToF) detector was exclusively considered for  $\pi/K/p$  separation in the SPD experiment [29]. Recently, the addition of an aerogel counter in the end-cap region was proposed to improve  $\pi/K$ -separation [13, 20]. This opened unique opportunities for machine learning applications as well. Machine learning has already been successfully applied to calibration and reconstruction of Cherenkov detectors [15].

Our thesis focuses on leveraging machine learning techniques to develop an end-to-end FARICH reconstruction model. We show that our model outperforms alternative multi-stage approaches, both algorithmic and machine learning based.

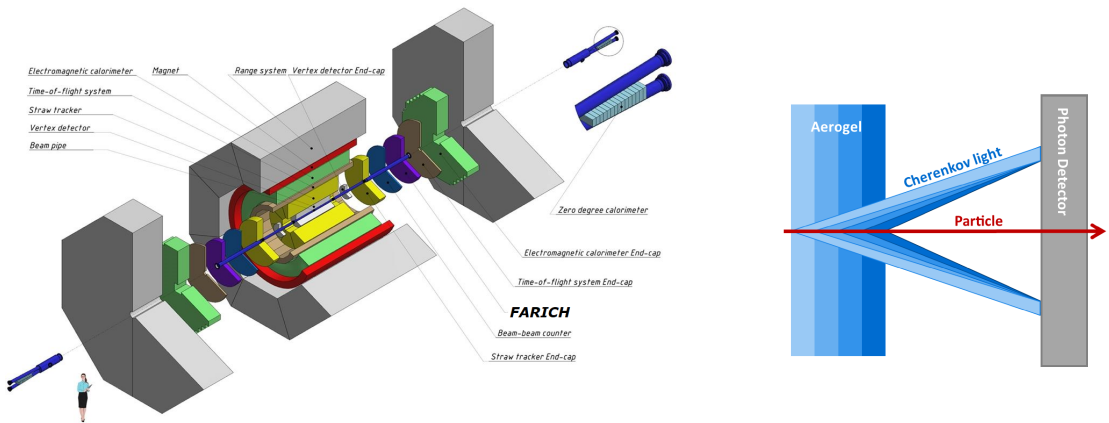


Figure 1.1: Left to right: FARICH in the SPD experiment [20], aerogel detector layout [8].

## Chapter 2

# Literature Review

Many of the machine learning approaches developed in the recent years have been incorporated in the HEP data processing. The SPD project considers utilizing machine learning for trigger system and event reconstruction in several stages of the data acquisition pipeline and in the data processing facility. Fast tracking is instrumental to the trigger system at the SPD. There is a rich history of successful applications of machine learning for these tasks. For example, boosted decision trees were implemented in FPGIs as a level-1 trigger system at the CMS experiment [2]. GEM tracking performance was substantially improved with neural networks [6, 24].

Computer simulations are one of the crucial components of HEP analysis. They provide a detailed theoretical reference for understanding the experimental data, against which models of both known and “new” physics can be tested. Nowadays, various surrogate models have been extensively adopted for this task, commonly Generative Adversarial Networks (GANs), e.g. in LHCb experiment electromagnetic calorimeter [26, 27, 11] and Belle II pixel vertex detector simulations [31]. Variational Autoencoders (VAEs) can also be used [30]. GANs with an auxiliary regressor help improve the reproducibility of marginal metrics [28]. GANs have been successfully applied for DIRC Cherenkov detector simulation in LHCb [14].

### 2.1 RICH PID

Common RICH particle identification (PID) methods are based on calculating various statistics of the Cherenkov angle distribution. RICH PID in the LHCb [25] and AMS [7] experiments was performed by discriminating mass hypotheses via a maximum likelihood estimation. Prior particle track info and detector response were used to determine Cherenkov angles for each photon hit and compute PID statistics. Belle II experiment adopted an approximation of the Cherenkov angle distribution with a Gaussian [19].

Recently, machine learning approaches have seen some use in the task of RICH reconstruction. Notably, DeepRICH used VAE to extract features from a raw detector output and a convolutional neural network (CNN) to classify the latent representations [17]. The detector of choice for DeepRICH was FastDIRC, an open source simulation package of internal reflection Cherenkov detectors. DeepRICH combined reconstruction and simulation in a single framework and benefited from bypassing low-level details of the system, however, it was not trained in an end-to-end fashion and did not use the hit time information.

The following work, DeeperRICH, expanded the method to PID of pions and kaons in the GlueX experiment at JLab [16]. In a DIRC detector, no distinct rings are observed as Cherenkov photons are internally reflected; an intricate hit pattern of several curved traces is formed instead. Specifics of the DIRC detector setup also include the use of a non-rectangular detector matrix. It is split into identical square patches, making it uniquely suitable for a Vision Transformer (ViT) model. The authors replaced the DeepRICH CNN with a Swin variant of vision transformer. It should be noted however that Swin is a notoriously computationally heavy model. DeeperRICH required multiple days of training on a cluster of A30 GPUs, moreover, its slow inference made real time processing challenging.

Recently, a machine learning based Gaussian mixture model (GMM) for predicting a prob-

ability density function of an algorithmic PID classifier at LHCb was explored [18]. Instead of relying only on simulations, the authors used real calibration data to better predict PID probabilities in complex feature spaces. A fully connected neural network was trained to model the dependence of GMM parameters on the set of experimental features. This approach allowed the model to smoothly interpolate over the feature space and account for non-trivial correlations between input features.

## 2.2 Current Analogous Methods

Moving on to the research efforts directly related to ours, a neural network for particle identification at the newly upgraded LHCb RICH detector was devised [9]. LHCb RICH is a traditional mirror-focusing detector with tractable geometry that produces full Cherenkov rings. The proposed method of obtaining images for CNN input encompassed a single polar transformation of a fixed radius range around each Cherenkov ring center. Similarly to our method, the center was defined as an intersection of the extrapolated particle track with the photon detection plane.

The polar transformation unwound the rings into vertical lines in 2d images. These images were then fed into a custom CNN that was trained in an end-to-end fashion to perform multi-class classification over 6 particle types. Cross entropy loss was used as a training objective. The CNN architecture used dropout for regularization and included 3 convolutional layers with increasing numbers of filters, as well as 2 final fully connected layers. The model was trained on a simulated dataset.

The proposed method was fine-tuned for the LHCb RICH detector. The detector only captures particles travelling at shallow angles of incidence, therefore, the resulting Cherenkov ellipses can not be as eccentric as in the SPD FARICH. The authors could skip implementing the full reconstruction of Cherenkov angles required for correct processing of highly eccentric rings and limit themselves to a simple non-parametric polar transformation. For similar reasons, they did not have to account for refraction of Cherenkov photons exiting the aerogel radiator. Accuracy was used as a metric during training as the simulated dataset was properly balanced.

## 2.3 Relation to Current State of the Art

Our methods build upon the contemporary machine learning based RICH PID approaches by utilizing end-to-end training for the classification of particle types. Similarly, our model trains on simulated physics data. These industry standard simulations are indispensable in HEP research, as they supply an extremely detailed and physics grounded description of the processes happening in the detector. The simulation software packages are built from the first principles, thus, provided the simulated detector configuration is well-matched with its physical implementation, the processes can be simulated as close to the real ones as possible [5].

Until the detector complex is fully realized in hardware, real data can be scarce. Nevertheless, the detector cannot be built without an already working data processing system, hence, despite its flaws, the simulated data provides a solid foundation for the development of reconstruction methods and, crucially, guides the process of hardware implementation towards matching the simulated setup more closely.

The methods described in this thesis markedly improve upon the current state of the art by implementing various elaborate Cherenkov ring processing algorithms with analytic corrections prior to feeding it to the neural network, adopting a modern residually connected CNN, employing sophisticated machine learning tricks to optimize the training, particularly for imbalanced datasets, monitoring advanced classification metrics during training, such as areas under various types of ROC curves and thresholded metrics, and comparing neural networks with several algorithmic



baselines, as well as with non end-to-end trained models. Additionally, we perform a comprehensive ablation study to dissect the real benefits of neural networks and other types of changes.

As our setup did not involve an internally reflected Cherenkov detector with an irregular SiPM matrix and considering the extremely high data rate of the SPD experiment, a heavy ViT model was deemed impractical, although it should definitely be considered for future research.

## 2.4 NICA SPD Details

The aerogel counter in the SPD project is expected to provide  $\pi/K$ -separation in the momenta range below 5 GeV/c, necessitating a focusing aerogel (FARICH) [13]. A robust reconstruction of events in FARICH is required, i.e. identifying the correspondence between tracks supplied by the online trigger and Cherenkov rings.

There are 2 possible cases: the use of silicon photo-multipliers (SiPMs) or microchannel plate detectors (MCPs). The SPD is characterized by long bunches up to 76 ns [13]. In the first case, this results in a significant background rate which should be countered with some kind of software algorithm, possibly employing machine learning techniques.

In the second case, the problem of superimposed Cherenkov rings arises. It is expected that the geometric coordinates of the Cherenkov cone vertices will be known with sufficient precision. It follows that the error will be localized in the time domain.

Unfortunately, the application of hardware reconstruction is not possible because of the tracking system delay. A purely software-based approach would need to find a matching between rings, tracks and vertices. It is argued that machine learning may help to speed up computations compared with a brute-force approach.

## Chapter 3

# Problem Statement

Ring-imaging Cherenkov (RICH) detectors use Cherenkov photons to measure the velocity of an elementary particle. Whenever a particle travels at a velocity higher than the speed of light in the medium, a Cherenkov radiation is emitted. Cherenkov photons form a cone that imprints a roughly elliptically shaped signal on a flat detector surface. A well-known relation between the velocity of a particle and the Cherenkov angle is as follows:

$$\beta = \frac{1}{n \cos \theta_c}, \quad (3.1)$$

where  $\beta$  denotes the particle velocity in units of  $c$ ,  $n$  is the refractive index of the medium, and  $\theta_c$  is the Cherenkov angle.

The main components of a RICH detector are radiator and photomultiplier (PM) array. Cherenkov cone vertex is not static and traverses the full thickness of the radiator during the passage of the particle, resulting in a smudged Cherenkov ring. Proximity-focusing (proximity-imaging, non-focusing) aerogel detectors compensate for this during the reconstruction stage [7]. However, in case a high resolution is desired, Cherenkov photons must be focused to reduce the signal variance. Focusing mirrors can be employed to reflect the photons back to the PM array, however, they are difficult to install, expensive to maintain, and require a large space that is at a premium in colliding beams experiments [4].

Aerogel has seen extensive use as a radiator material for as soon as high quality samples appeared on the market. The main idea behind the Focusing Aerogel RICH (FARICH) detector is stacking several layers of aerogel with increasing refractive index in the radiator (Fig. 1.1, right). The focusing happens inside the radiator, eliminating the need for mirrors, drastically reducing the mechanical complexity [19].

### 3.1 FARICH Reconstruction

In the task of FARICH reconstruction, one is generally provided with the following input:

- Track parameters:  $x_p, y_p, z_p$  – coordinates of the particle upon entering the aerogel,  $\theta_p, \phi_p$  – polar angle and azimuth of the direction of travel;
- Photon hits:  $x_c, y_c, z_c$  – coordinates of triggered pixels in the photosensitive matrix.

SiPMs used in the matrix can be triggered by Cherenkov photons, as well as background radiation and intrinsic noise.

The end goal of PID is to identify the most probable particle, however, it is not possible without knowing the momentum  $p$ . Therefore, one should extract as much useful statistics from the data as possible, such that when combined with an estimated momentum value from a separate detector, the particle is robustly identified. This can be formally written as an optimization problem

$$\mathbb{E}_{(X,y) \sim q(X,y)} [Q(f_W(X), y)] \rightarrow \max_W,$$

where  $f_W$  is a classifier model with parameters  $W$ ,  $X$  is the input described above including momentum,  $y$  is a particle class,  $q$  is a distribution of detector responses, and  $Q$  is a classification metric, e.g. accuracy or  $F_1$ -score. In the context of HEP, individual probabilities of misidentification or even full confusion matrices are usually of interest.

A simple example of  $f_W$  is to estimate the velocity  $\hat{\beta}$  and compute a mass hypothesis:

$$\hat{m}_0 = \frac{p}{\gamma \hat{\beta}}, \quad \gamma = \frac{1}{\sqrt{1 - \hat{\beta}^2}}.$$

Estimator  $\hat{\beta}$  can be either algorithmic or learnable. In the former case,  $\theta_c$  is explicitly calculated for each photon hit. A statistic  $\hat{\theta}_c$  can then be derived from this sample. Model parameters  $W$ , e.g.  $\theta_c$  threshold, are hyperparameters and not learnable in this case. In the latter case, a neural network or other model may be trained to predict  $\hat{\beta}$ , ultimately solving a regression task. The momentum  $p$  is used in the next stage to calculate the rest mass.

## 3.2 End-to-End Optimization Proposal

We propose a different way of constructing  $f_W$  by encapsulating all stages of data processing in a single learnable function that has access to FARICH, tracking and momentum data. The potential benefit of this is two-fold.

Firstly, the  $\beta$  distribution is inherently imbalanced as a consequence of conservation of momentum. Lighter particles, such as electrons, obtain ultra-relativistic speeds and concentrate near  $\beta = 1$ , while heavier particles come in a broader range of lower velocities, making the regression task harder to optimize.

Secondly, during training an error in the intermediate value of  $\hat{\beta}$  is minimized. A proxy objective then implicitly influences the target classification metric  $Q$ . This may lead to issues, for example, improving margins on  $\beta$  in cases where the particle type can be predicted with high certainty does not improve PID performance. The training should instead focus on hard events where classification typically fails.

This work investigates the potential benefits of an end-to-end model over a multi-stage model. For this purpose, we develop both types of models and test them. We consider simple algorithmic reconstruction methods, a neural network regressor for  $\beta$ , and various end-to-end neural networks. The models are evaluated on synthetic datasets and tested against several baselines to find the best performing approach which could be used as a basis for future development of FARICH reconstruction.

## Chapter 4

# Methods

Our first baseline was based on the RICH reconstruction from CBM experiment at FAIR [21]. The algorithm utilized Hough transform for ellipses, more precisely the Taubin method [12]. The Hough transform estimated parameters of the ellipse, such as  $a, b$  – semi-axes,  $x, y$  – center coordinates, and  $\phi$  – angle of rotation (Fig. 4.1). These parameters were then used to calculate  $\theta_c$  (3.1).

It is important to note that the baseline was fixed, so we only had access to the precomputed  $\theta_c$  values. The baseline did not account for refraction in the aerogel. More importantly, it did not have access to information from the straw tracker, such as  $\theta_p$  and  $\phi_p$ .

Similarly to our own algorithmic methods, the second fixed baseline used Cherenkov angle distribution (4.1):

$$\theta_c(\phi_c|\beta, n, \theta_t) = \arccos\left(\frac{1}{n\beta}\right) + \arccos\left(n(1 - (\mathbf{n}_0\mathbf{n}_\gamma)^2)\right) + (\mathbf{n}_0\mathbf{n}_\gamma)\sqrt{1 - n^2(1 - (\mathbf{n}_0\mathbf{n}_\gamma)^2)} \quad (4.1)$$

A regression fit was computed over real Cherenkov hits with this formulae. The method accounted for refraction and utilized tracking data. It was used in the second dataset, however, its implementation contained critical errors and produced garbage output for 50% of the data. We removed cases where the algorithm malfunctioned from the test set for the sake of comparison. The reader is advised to keep in mind that production baseline was evaluated on a subset of the test sample and exercise caution when drawing conclusions in relation to other baselines.

### 4.1 Refraction Correction

Cherenkov photons are refracted upon exiting the radiator (Fig. 4.2), making a direct computation of  $\theta_c$  from  $x_c, y_c$  inaccurate. One possible way to account for this is to reverse the refraction and obtain initial  $\theta_c$  values.

Let  $\alpha$  be the angle of incidence,  $\beta$  – the angle of reflection,  $n$  – refractive index of the

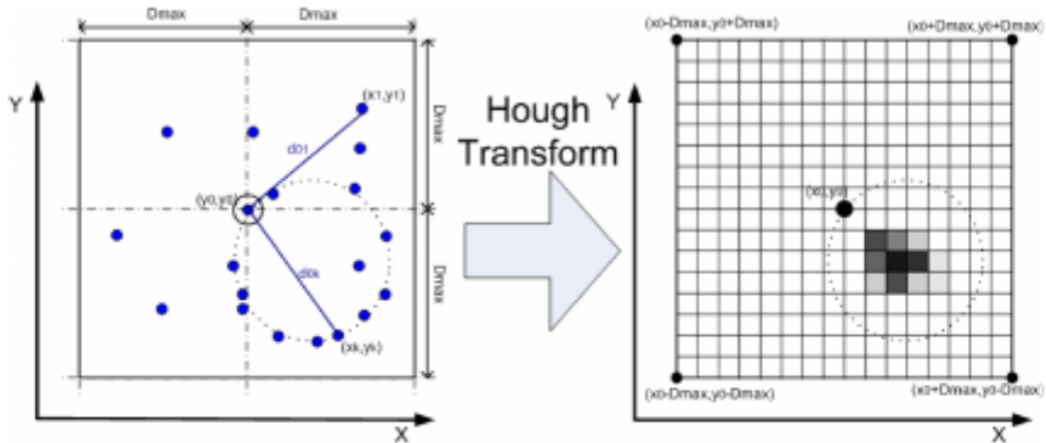


Figure 4.1: RICH Hough transform baseline [21].



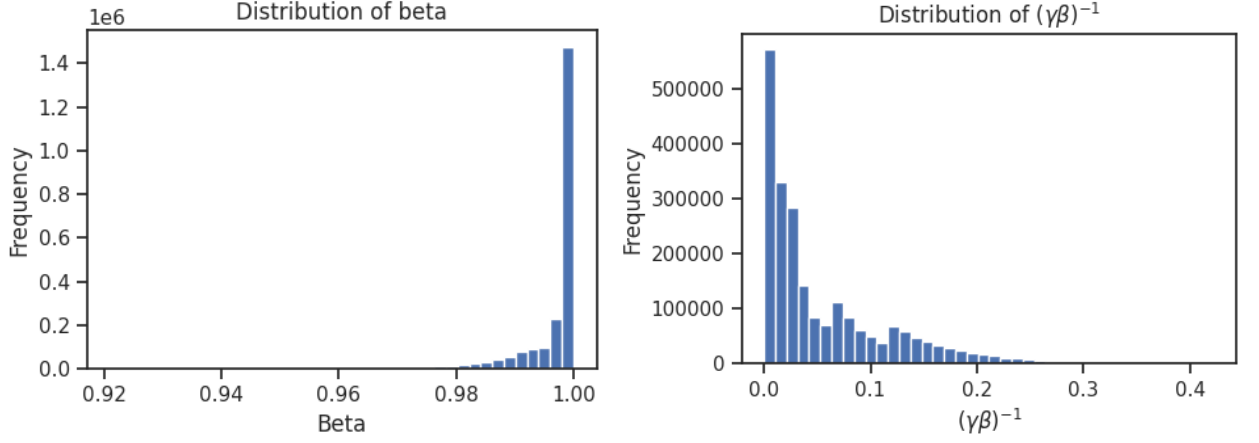


Figure 4.4:  $\beta$  distribution in the data. The sample is heavily skewed because of the relativistic electrons produced in collisions.

## 4.2 Algorithmic Methods

$\theta_c$  values can be used to estimate  $\hat{\theta}_c$ . Several approaches are possible here.  $\hat{\theta}_c$  can be set to a simple arithmetic mean. This is equivalent to  $l_2$  optimization in the regression fit of a production baseline. Median or  $l_1$  method computes the median of  $\theta_c$  after correction. Maximum likelihood (MLE) finds an approximate mode of the distribution using a sliding window approach. We implemented the latter 2 approaches. An arithmetic mean was supplied by a production baseline.

## 4.3 Machine Learning Models

Algorithmic methods output  $\hat{\theta}_c$  that can be converted to particle rest mass using momentum  $p$  from the straw tracker. In machine learning terms, these are regression models. However, the goal of RICH detector is separating different particle types. In particular, FARICH aims to improve  $\pi/K$ -separation. This can be stated as a classification problem.

Unlike algorithmic methods, end-to-end models are able to skip the intermediate step of computing velocity  $\hat{\beta}$ . The regression task is not well suited for a machine learning model because of the significant value imbalance stemming from the high velocities of light particles, e.g. electrons (Fig. 4.4, left). A mitigation in the form of compensating  $\beta$  with Lorentz factor  $\gamma$  can be applied, however, the distribution will still be far from a properly balanced one (Fig. 4.4, right).

Taking this into account, we propose a neural classifier that combines hit coordinates  $x_c, y_c$ , tracking info  $x_p, y_p, z_p, \theta_p, \phi_p$  and momentum  $p$  in a single architecture. Several approaches are

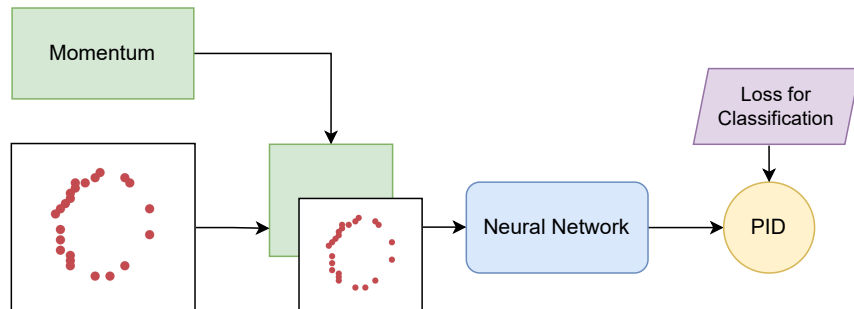


Figure 4.5: Extra Channel NN architecture.

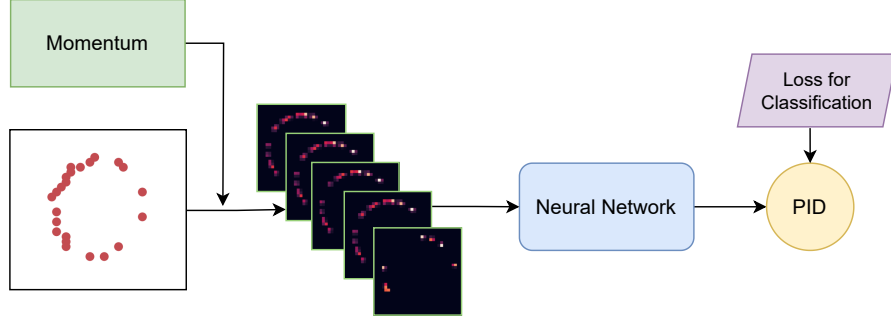


Figure 4.6: Renormalize NN architecture.

possible: either providing  $p$  as an extra input channel / pathway in the neural network, or employing it in the input data normalization. For the sake of completeness, we implemented both methods.

In the first case, the normalized momentum value is supplied as an additional input channel, hence the name *Extra Channel NN*. Hit coordinates  $x_c, y_c$  are geometrically transformed into a circular response using the particle track information (Fig. 4.5). This way, the network learns a simpler task of extracting the radius of a circle instead of dealing with diverse elliptical responses of various orientations and eccentricities. The model is also supplied with 2 extra channels of Fourier features computed from  $\phi_p$  values.

The second model, *Renormalize NN*, deals with momentum by transforming the input circle even further. The non-linear dependence on  $p$  can be disentangled and removed prior to applying classifier model. The following relation stands:

$$\cos \theta_c = \frac{1}{n} \sqrt{m_0^2/p^2 + 1}, \quad (4.3)$$

where  $m_0$  is the rest mass.

Renormalize NN first corrects for refraction, then uses (4.3) with  $N$  mass probes for each of the particle types in the data to transform  $\theta_c$  into a  $\beta$ -independent space. The input is then split into  $N$  channels. Coordinates  $x_c, y_c$  are recomputed using corresponding mass probes (Fig. 4.6).

It can be easily seen that multiplying  $\cos \theta_c$  by an inverse of (4.3) results in  $\cos \theta_c = 1$  in case  $m_0$  is the true rest mass of the particle, and  $\cos \theta_c \neq 1$  otherwise. This way, a Cherenkov ring of a fixed size is placed in the channel corresponding to a correct particle type, while rings of different sizes are placed in other channels.

In theory, by thresholding around a known fixed radius, all channels besides the correct one can be cleared of the hits entirely. In practice, the threshold can not be set too small because of the finite thickness of the Cherenkov ring. We tune the threshold value such that it eliminates at least some of the hits from the wrong probes and does not remove too much relevant hits from the correct mass probe.

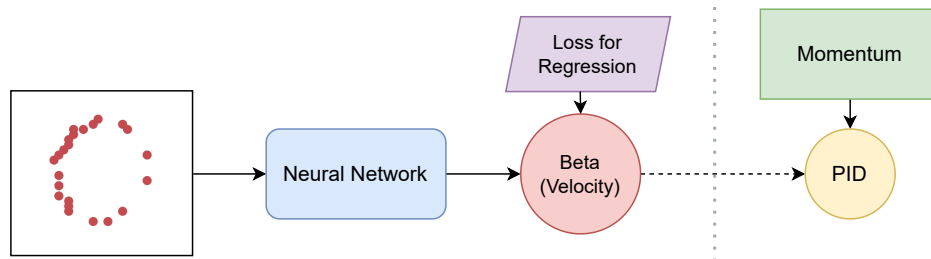


Figure 4.7: Beta NN architecture.

The described normalization approach may relieve the model from implicitly learning the non-linear  $\beta \leftrightarrow p$  relationship, enabling a more efficient use of its computational power. Potential drawbacks include the loss of information during thresholding and the introduction of additional granularity in the momentum value  $p$ .

Additionally, we implement a  $\beta$ -outputting neural regressor *Beta NN* (Fig. 4.7). It is a direct analog of algorithmic methods with the only difference being the use of a neural backbone to perform  $\hat{\beta}$  computation. The model learns to predict  $\beta$  and then uses (3.1) to classify the particle.



## Chapter 5

# Numerical Experiments

Two distinct datasets were utilized in the numerical experiments: Balanced and Production. Both were produced in the SPDRoot particle passage simulation toolkit [13]. The data samples contained 5366595 and 27871018 simulated events respectively, consisted solely of Cherenkov photons, and did not include background hits and scattered photons. High quality photomultipliers are expected in the NICA SPD FARICH, so the real background rate should be fairly low.

Balanced dataset was well-balanced with 5 particles present:  $e^-$ ,  $\mu^-$ ,  $\pi^+$ ,  $K^+$ , and  $p$ . Beam collisions were not simulated, so the particles were assigned a random momentum drawn from a uniform distribution.

Balanced dataset contained events with low  $\beta$  values that did not produce any Cherenkov photons (Fig. 4.4). There were also events with  $\theta_p > \pi/2$  meaning particles were moving backwards through the detector. We excluded these events from our analysis because they led to a breakdown of all reconstruction algorithms. In total, such events accounted for a 1.7% of the dataset. Production dataset did not exhibit such issues, but still contained rare events with zero proper Cherenkov photons as they were all scattered.

Production dataset used a realistic FARICH geometry with 2 end-cap detectors. Beam collisions were properly simulated resulting in realistic distributions of parameters (Fig. 5.1). Because of the significant class imbalance ( $\pi/K$  ratio  $\approx 20/1$ ), we focused our efforts on  $\pi/K$  separation only. To counter class imbalance, we enabled either a class-aware downsampling or a positive weighting of the loss function. A control run without applying a balancing method was also performed for each model. We compared all 3 approaches by calculating AUROC on the validation set and used the best performing checkpoint for the test set evaluation.

We reserved 600000 events for testing, 5000 for validation, and used the remaining for model training and calibration. The input format for NN models was a  $C \times 32 \times 32$  tensor, containing either a momentum value, a Cherenkov ring and its Fourier features or mass probes for corresponding particles. An adapted ResNet-18 architecture was used as a classifier and regressor with changes to the input convolution, max pooling and a classifier head / sigmoid activation to

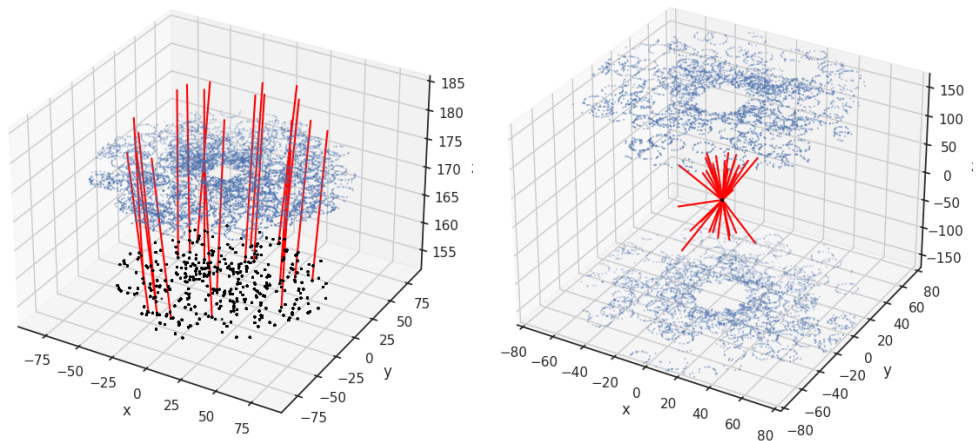


Figure 5.1: FARICH configurations in the simulated data. Left – Balanced, right – Production.

accommodate for input and output formats.

All models were trained on  $\approx 3100000$  events with Adam optimizer, cosine annealing scheduler, a learning rate of  $5 \cdot 10^{-4}$  and a batch size of 128. Cross entropy loss was optimized for classification models, while Beta NN was trained with MSE loss. Multiclass accuracy, precision, recall and AUROC were monitored during training. For Beta NN, RMSE and R2 were also logged.

In total, 2 end-to-end models and 4  $\beta$ -outputting baselines were considered: Extra Channel NN, Renormalize NN, learnable Beta NN and algorithmic Median, MLE and Hough / Fit baselines.

# Chapter 6

## Results and Discussion

### 6.1 Regression Metrics

We used regression metrics to tune hyperparameters of the algorithmic methods. A sample standard deviation  $\sigma_\beta$  was our main target metric. We noticed that the refraction correction method, while improving the variance, introduced a systematic bias in the predictions. Fortunately, the bias could be approximated by a linear transformation quite well, and thus easily corrected. Systematic correction parameters were computed by making a single pass over a subset of the training data and then applied to predictions on a test set.

This improved  $\sigma_\beta$  by more than 40% for Median on Balanced dataset. An improvement of 10% was achieved on Production dataset. MLE estimate also improved, however, Hough baseline standard deviation increased. A possible reason for this can be found in the supplementary material (Appendix A: Ablation Study), where we demonstrate the improvement of classification metrics after correction. To this end, we used corrected values in the subsequent analysis.

Among our  $\beta$ -outputting baselines, Beta NN showcased the best variance  $\sigma_\beta$ , although the advantage over the second best performing Median baseline was small on Balanced dataset (Fig. 6.1, top). The wavy shape of the plot indicates a bias variance trade off encountered in the training process, making the model questionable from the physical validity standpoint.

The advantage was more pronounced on Production dataset (Fig. 6.1, bottom). In that case, the model remained physically sound across the whole  $\beta$  range. MLE and Hough / Fit baseline

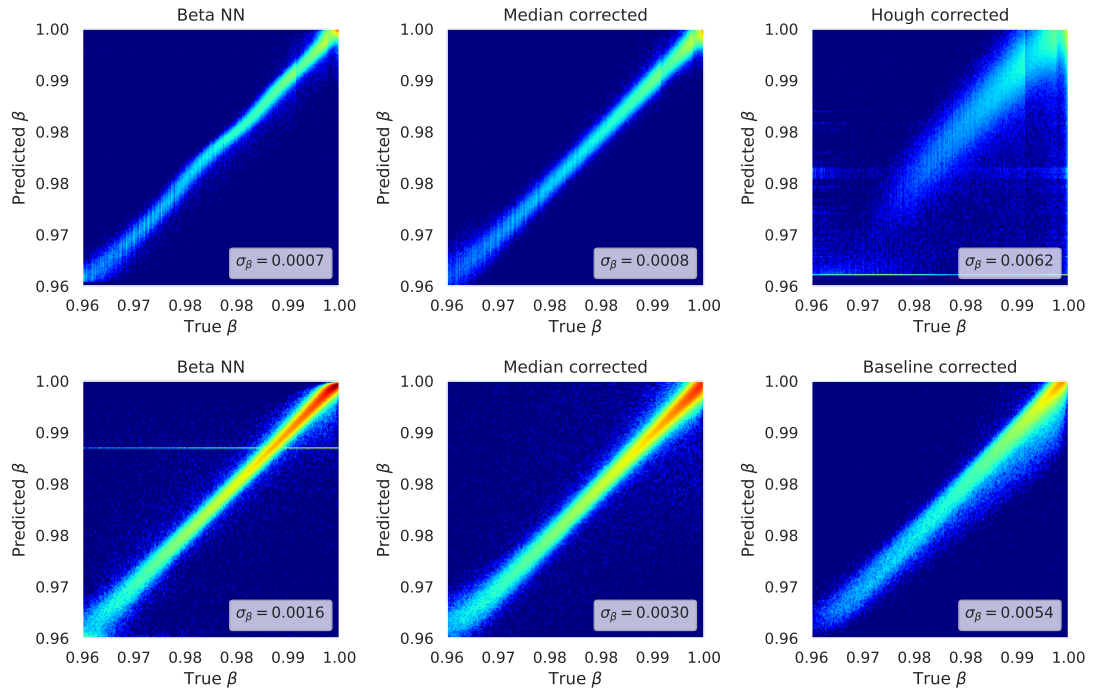


Figure 6.1: 2d correlation plots for  $\beta$ -outputting models. Top – Balanced, bottom – Production dataset.

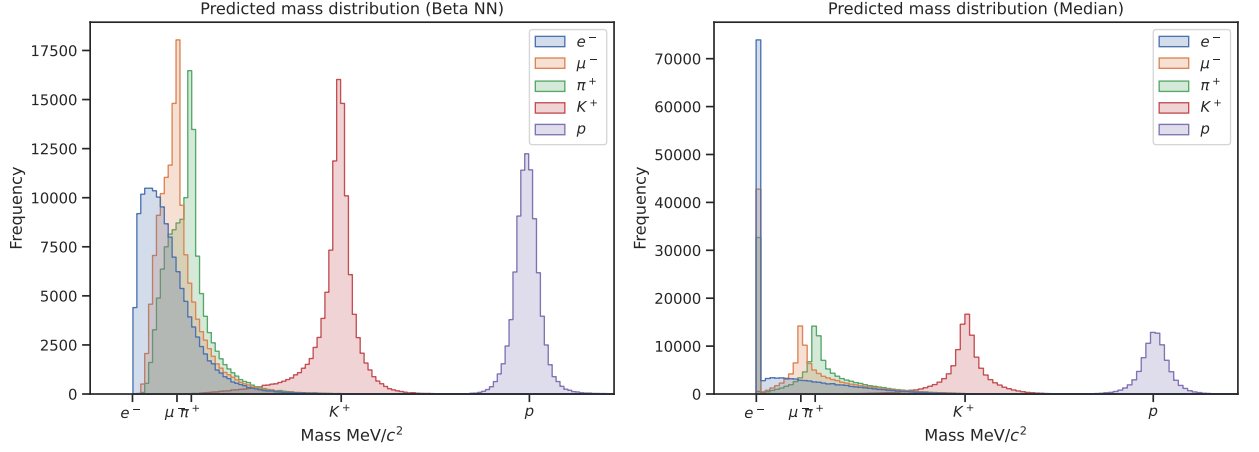


Figure 6.2: Predicted mass distribution, color coded by true classes.

performed less favorably (Table 6.1).

An interesting artifact can be observed in the Beta NN 2d correlation plot on Production dataset. The horizontal line corresponds to events with no proper Cherenkov photons registered. The model learned to guess an average  $\beta$  value in such cases. In the future, this behavior can be made more explicit by adding an additional class for the unknown particle type.

## 6.2 Classification Metrics

We computed rest masses from the predictions of  $\beta$ -outputting models and used statistical hypotheses to determine particle types. The hypotheses were based on the proximities of predicted masses to true particle rest masses. For end-to-end models, NN output after softmax supplied probabilities and predicted classes directly.

We considered the following classification metrics: area under receiver operating characteristic curve (AUROC), precision at a fixed recall value (Purity @ Efficiency = 0.99), average precision or AUC-PR (AP), and 5 class accuracy for Balanced data.

End-to-end NN models held a significant advantage over all  $\beta$ -based approaches, both the best algorithmic model and Beta NN (Table 6.1). Median demonstrated the best performance among all algorithmic methods, nevertheless, it was outperformed by NN models. Extra Channel NN showed the best overall performance on both datasets (Table 6.2).

The 5 class accuracy of all methods, including NN, was not very high because lower mass particles were not separated well (Fig. 6.2). This is not an indicator of a bad performance. The minimal amount of particle band overlap is determined by the detector configuration, thus limiting the accuracy even in the case of an ideal theoretical algorithm (Fig. 6.3). The bands possess a

Table 6.1: Test metrics for Balanced dataset. AP stands for Average Precision (AUC-PR). For  $\pi/K$  metrics, a subset of the test data containing only pions and kaons was taken.

Method	5 class Accuracy	$\pi, K$ Purity@Efficiency = 0.99	$\pi, K$ AP	$\sigma_\beta$
<b>Extra Channel NN</b>	<b>0.68</b>	<b>0.946</b>	<b>0.998</b>	N/A
Renormalize NN	0.67	0.946	0.998	N/A
Beta NN	0.66	0.945	0.992	0.0007
Median	0.65	0.827	0.985	0.0009
MLE	0.53	0.467	0.828	0.0020
Hough	0.50	0.461	0.719	0.0062

Table 6.2: Test metrics for Production dataset. AP stands for Average Precision (AUC-PR).

Method	AUROC	Purity@Efficiency = 0.99	AP	$\sigma_\beta$
<b>Extra Channel NN</b>	<b>0.9999</b>	<b>0.93</b>	<b>0.998</b>	N/A
Renormalize NN	0.9997	0.81	0.994	N/A
Beta NN	0.9990	0.91	0.937	0.0016
Median	0.986	0.05	0.867	0.0030
MLE	0.977	0.05	0.713	0.0038
Baseline	0.917	0.05	0.886	0.0054

finite spread, making particles physically indistinguishable at higher momenta. We observe this in practice by plotting classification metrics binned by momentum (Appendix B: Additional Figures).

The primary goal was  $\pi/K$ -separation, where the overlap is not as severe as seen in muons and electrons, reducing the impact of total accuracy. In particular, for an important momentum range of  $5.0 \pm 0.5$  GeV/ $c$  and an efficiency fixed at 0.99, the best NN model improved  $\pi/K$  purity by almost 20% over the best algorithmic method.

Notably, on Balanced dataset Extra Channel and Renormalize NN differed in multiclass accuracy only. They produced identical outputs on  $\pi/K$  subset. These NNs received vastly different inputs, indicating that their performance is close to optimal for this particular dataset.

### 6.3 Discussion

Our  $\beta$ -outputting methods surpassed the baselines because we carefully accounted for refraction and introduced strict physical thresholds on  $\theta_c$  values, resulting in sharper  $\beta$  distributions. Hough transform’s disadvantage was its high dimensionality in the case of ellipses (5 parameters) in addition to the dismissal of tracking information. While the transform worked somewhat reliably for simple circles, it faced challenges when locating ellipses. Production baseline was more advanced, however, in that case all 3 algorithmic methods suffered from poor performance at classifier thresholds corresponding to high efficiencies (Table 6.2).

Beta NN performance was not quantitatively different from algorithmic methods on Balanced dataset. Qualitatively, it exhibited unfavorable deviations from proper physics. Beta NN performed well on Production data because it processed empty events more efficiently. Algorithmic methods outputted  $\beta = 1$  in such cases. Thus far, an argument can be made in favor of our initial hypothesis that most of the quality uplift in NNs is explained by the end-to-end training.

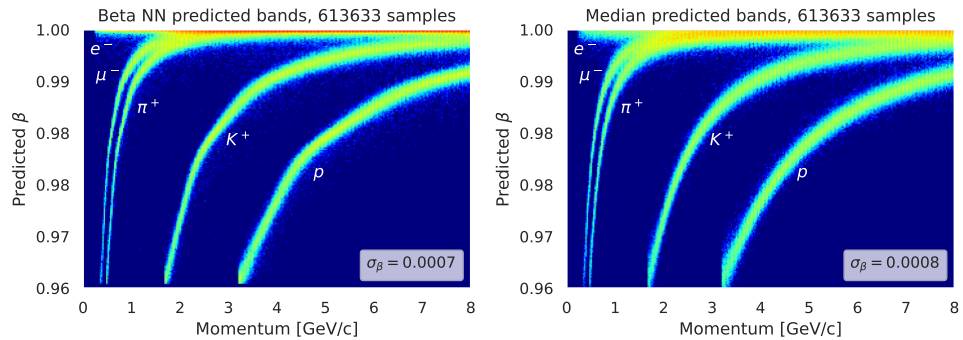


Figure 6.3: Predicted bands on Balanced dataset. Some of the overlap in the bands stems from inherent physical limitations and can never be completely removed, making 100% accurate classification impossible.

Interestingly, Purity @ Efficiency of Beta NN was higher than that of an end-to-end Renormalize NN. All the integral metrics were lower however, hinting at a potential issue with Renormalize NN in a particular narrow segment of the precision recall curve. This may be investigated further.

Summing up, the excellent quality of end-to-end NN models can be reasonably attributed to their training pipeline. The same NN architecture that was not trained in an end-to-end fashion did not demonstrate significant advantages over algorithmic methods, despite having access to additional spatial information that was not available to algorithmic approaches. This underlines the benefits of optimizing a classification objective rather than an intermediate regression metric.

While neural networks demonstrated impressive capabilities, the use of ResNet-18 CNN in the production setting of the SPD data processing facility is challenging. In the future work, one may optimize the architecture with various techniques such as knowledge distillation, pruning, and quantization. A multi layer perceptron (MLP) may serve as a target architecture as it is more suited for running on CPUs.

## Chapter 7

# Conclusions

ML-based approaches have gained attention of physicists across many different research fields. In particular, applications in particle and nuclear physics, ranging from experiment design to fast experimental data processing, offer great prospects for enabling new scientific discoveries [10]. Deep learning applications have been recently explored in the context of fast and reliable reconstruction and simulation of RICH detectors. In particular, neural networks designed for fast reconstruction offer higher accuracy and robustness compared to traditional methods.

The FARICH detector deployment at NICA SPD gives rise to a range of essential problems: fast online noise rejection, offline reconstruction, fast simulation, and more. These tasks exhibit a high potential for ML applications. ML proved useful across a wide range of HEP applications, in particular, RICH detectors, hence, applying ML to FARICH poses a great opportunity.

In this work, we developed and tested single ring reconstruction methods for NICA SPD FARICH. The initial results of applying ML on Balanced dataset were promising, with a lot of useful implications and potential for future studies. They were published in Moscow University Physics Bulletin journal. The journal is peer-reviewed, published in English by Springer and indexed by Scopus and WoS. Production sample results were presented at IX SPD Collaboration Meeting.

The logical next step would be to finalize the SPD reconstruction method into a production-ready solution. This involves exporting the neural network code from Python environment to C++ and integrating it into the SPDRoot system. These steps were discussed at the recent SPD Collaboration Meeting.

Previously, FARICH was considered for SCTF project in Novosibirsk, Russia. Unfortunately, the progress of FARICH deployment has slowed down quite seriously after the collider was repurposed for a hard x-ray material science program. Recently, the prospect of collaborating on FARICH with SCTF China facility has been discussed. Our colleagues show interest in our expertise on the topic. This collaboration opens a world of possibilities to develop new reconstruction and simulation methods.

Now that FARICH has been officially adopted for the second phase of NICA SPD experiment, the applicability of our research is broader than ever. A substantial progress has already been achieved, with prospects wide open for further research, new exciting developments and partnerships.

## Chapter 8

# Author Contribution

Among the 4 co-authors of this work, the student's contribution was the development and implementation of machine learning methods and baselines, data processing, numerical experiments, model evaluation and analysis, as well as composing the results, presenting at conferences, writing papers and reviewing the relevant literature.

Ivanov, A. contributed simulated datasets and 2 baselines: Hough transform (Balanced) and Fit (Production).

Barnyakov, A. and Ratnikov, F. contributed the overall direction and supervision, guidance in technical physics topics and final editing of the written works and conference talks.



## Chapter 9

# List of publications

The student contributed the development of the main machine learning method and baselines to all of the papers, with the exception of paper 1 which had 2 baselines provided by Ivanov, A. The co-authors contributed the simulated Geant4 / SPDRoot datasets, important physics details for the detector configuration and overall supervision.

Paper 1 presents the results obtained on Balanced dataset in the first year of MSc studies. The production dataset results were reported at IX SPD Collaboration Meeting (2025 May 12-16).

Papers 2 and 3 are concerned with FARICH reconstruction and noise suppression for Super  $c - \tau$  Factory experiment instead of NICA SPD. Nevertheless, they laid an important groundwork for this thesis. However, the methods developed during the MSc studies are significantly novel. They were not present in the previous BSc papers in any way, because the task of an end-to-end particle identification was never examined.

1. Shipilov, F., Barnyakov, A., Ivanov, A., Ratnikov, F. Machine Learning for FARICH Reconstruction at NICA SPD. In *Moscow University Physics Bulletin* (2024), vol. 79, no. S2, S906–S913.
2. Shipilov, F., Barnyakov, A., Bobrovnikov, V., Kononov, S., Ratnikov, F. What Machine Learning Can Do for Focusing Aerogel Detectors. In *EPJ Web of Conferences* (2024), vol. 295, 09043.
3. Shipilov, F., Barnyakov, A., Bobrovnikov, V., Kononov, S., Ratnikov, F. What Machine Learning Can Do for Focusing Aerogel Detectors. In *Physics of Atomic Nuclei* (2023), vol. 86, no. 5, 864–868.

# **Acknowledgements**

The research leading to these results has received funding from the Basic Research Program at the National Research University Higher School of Economics.

# Bibliography

- [1] Abazov, V., Abramov, V., Afanasyev, L., Akhunzyanov, R., Akindinov, A., Akopov, N., Alekseev, I., Aleshko, A., Alexakhin, V. Y., Alexeev, G., et al. Conceptual design of the spin physics detector. *arXiv preprint arXiv:2102.00442* (2021).
- [2] Acosta, D., Brinkerhoff, A., Busch, E., Carnes, A., Furic, I., Gleyzer, S., Kotov, K., Low, J. F., Madorsky, A., Rorie, J., Scurlock, B., Shi, W., and on behalf of the CMS Collaboration. Boosted decision trees in the level-1 muon endcap trigger at cms. *Journal of Physics: Conference Series* 1085, 4 (sep 2018), 042042.
- [3] Albertsson, K., Altoe, P., Anderson, D., Andrews, M., Araque Espinosa, J. P., Aurisano, A., Basara, L., Bevan, A., Bhimji, W., Bonacorsi, D., et al. Machine learning in high energy physics community white paper. In *Journal of Physics: Conference Series* (2018), vol. 1085, IOP Publishing, p. 022008.
- [4] Alexeev, M., Birsá, R., Bradamante, F., Bressan, A., Chiosso, M., Ciliberti, P., Dalla Torre, S., Denisov, O., Duic, V., Ferrero, A., et al. Mirror alignment control for compass rich-1 detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 639, 1 (2011), 219–221.
- [5] Allison, J., Amako, K., Apostolakis, J., Araujo, H., Dubois, P. A., Asai, M., Barrand, G., Capra, R., Chauvie, S., Chytrcek, R., et al. Geant4 developments and applications. *IEEE Transactions on nuclear science* 53, 1 (2006), 270–278.
- [6] Baranov, D., Mitsyn, S., Goncharov, P., and Ososkov, G. The particle track reconstruction based on deep neural networks. In *EPJ Web of Conferences* (2019), vol. 214, EDP Sciences, p. 06018.
- [7] Barao, F., Arruda, L., Borges, J., Goncalves, P., Pimenta, M., and Perez, I. Cherenkov angle and charge reconstruction with the rich detector of the ams experiment. *Nucl. Instrum. Methods A* 502, 1 (2003), 310–314. In *Proceedings of the Fourth International Workshop on Ring Imaging Cherenkov Detectors*.
- [8] Barnyakov, A., and Collab., S. C.-T. F. The project of the super charm-tau factory in novosibirsk. *J. Phys.: Conf. Ser.* 1561, 1 (jun 2020), 012004.
- [9] Blago, M. P. Deep Learning Particle Identification in LHCb RICH. 2021.
- [10] Boehnlein, A., Diefenthaler, M., Sato, N., Schram, M., Ziegler, V., Fanelli, C., Hjorth-Jensen, M., Horn, T., Kuchera, M. P., Lee, D., Nazarewicz, W., Ostroumov, P., Orginos, K., Poon, A., Wang, X.-N., Scheinker, A., Smith, M. S., and Pang, L.-G. Machine learning in nuclear physics. *Rev. Mod. Phys.* 94, 3 (sep 2022), 031003.
- [11] Chekalina, V., Orlova, E., Ratnikov, F., Ulyanov, D., Ustyuzhanin, A., and Zakharov, E. Generative models for fast calorimeter simulation: the lhcb case. In *EPJ Web of Conferences* (2019), vol. 214, EDP Sciences, p. 02034.
- [12] Chernov, N. On the convergence of fitting algorithms in computer vision. *Journal of Mathematical Imaging and Vision* 27 (2007), 231–239.

- [13] Collaboration, T. S. Technical design report of the spin physics detector. JINR SPD Collaboration, 2023.
- [14] Derkach, D., Kazeev, N., Ratnikov, F., Ustyuzhanin, A., and Volokhova, A. Cherenkov detectors fast simulation using neural networks. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 952 (2020), 161804.
- [15] Fanelli, C. Machine learning for imaging cherenkov detectors. *JINST* 15, 02 (feb 2020), C02012–C02012.
- [16] Fanelli, C., Giroux, J., and Stevens, J. Deep (er) reconstruction of imaging cherenkov detectors with swin transformers and normalizing flow models. *Machine Learning: Science and Technology* 6, 1 (2025), 015028.
- [17] Fanelli, C., and Pomponi, J. Deeprich: learning deeply cherenkov detectors. *Machine Learning: Science and Technology* 1, 1 (2020), 015010.
- [18] Graziani, G., Anderlini, L., Mariani, S., Franzoso, E., Pappalardo, L., and di Nezza, P. A neural-network-defined gaussian mixture model for particle identification applied to the lhcb fixed-target programme. *Journal of Instrumentation* 17, 02 (Feb. 2022), P02018.
- [19] Kindo, H., Adachi, I., Burmistrov, L., Le Diberder, F., Hataya, K., Kakimimoto, S., Kakuno, H., Kawai, H., Kawasaki, T., Konno, T., Korpar, S., Križan, P., Kumita, T., Lai, Y., Machida, M., Mrvar, M., Nishida, S., Noguchi, K., Ogawa, K., Ogawa, S., Pestotnik, R., ?antelj, L., Sumiyoshi, T., Tabata, M., Tamechika, S., Yonenaga, M., Yoshizawa, M., and Yusa, Y. Initial performance of the aerogel rich detector of the belle ii experiment. *Nucl. Instrum. Methods A* 952 (2020), 162252. in *10th International Workshop on Ring Imaging Cherenkov Detectors (RICH 2018)*.
- [20] Korzenev, A. Spin physics detector at the nica accelerator complex. Technology & Instrumentation in Particle Physics (TIPP2023), 2023.
- [21] Lebedev, S., and Ososkov, G. Fast algorithms for ring recognition and electron identification in the cbm rich detector. *Physics of particles and nuclei letters* 6 (2009), 161–176.
- [22] Mikirtychyants, M., Engels, R., Grigoryev, K., Vasilyev, A., Chiladze, D., Kacharava, A., Mikirtychyants, S., Rathmann, F., Sarkadi, J., Schleichert, R., et al. First experiments with the polarized internal gas target at anke/cosy. In *Journal of Physics: Conference Series* (2011), vol. 295, IOP Publishing, p. 012148.
- [23] Müller, B., and Nagle, J. L. Results from the relativistic heavy ion collider. *Annu. Rev. Nucl. Part. Sci.* 56 (2006), 93–135.
- [24] Ososkov, G. A., Bakina, O. V., Baranov, D. A., Goncharov, P. V., Denisenko, I. I., Zhemchugov, A., Nefedov, Y. A., Nechaevskiy, A. V., Nikolskaya, A. N., Shchhavelev, E. M., et al. Tracking on the besiii cgem inner detector using deep learning. *Computer research and modeling* 12, 6 (2020), 1361–1381.
- [25] Powell, A. Reconstruction and pid performance of the lhcb rich detectors. *Nucl. Instrum. Methods A* 639, 1 (2011), 260–263. in *Proceedings of the Seventh International Workshop on Ring Imaging Cherenkov Detectors*.
- [26] Ratnikov, F. Generative adversarial networks for lhcb fast simulation. In *EPJ Web of Conferences* (2020), vol. 245, EDP Sciences, p. 02026.

- [27] Ratnikov, F., and Rogachev, A. Fast simulation of the electromagnetic calorimeter response using self-attention generative adversarial networks. In *EPJ Web of Conferences* (2021), vol. 251, EDP Sciences, p. 03043.
- [28] Rogachev, A., and Ratnikov, F. Gan with an auxiliary regressor for the fast simulation of the electromagnetic calorimeter response. In *Journal of Physics: Conference Series* (2023), vol. 2438, IOP Publishing, p. 012086.
- [29] Savin, I., Efremov, A., Peshekhonov, D., Kovalenko, A., Teryaev, O., Shevchenko, O., Nagajcev, A., Guskov, A., Kukhtin, V., and Toplilin, N. Spin physics experiments at nica-spd with polarized proton and deuteron beams. In *EPJ Web of Conferences* (2015), vol. 85, EDP Sciences, p. 02039.
- [30] Sergeev, F., Jain, N., Knunyants, I., Kostenkov, G., and Trofimova, E. Fast simulation of the lhcb electromagnetic calorimeter response using vaes and gans. In *Journal of Physics: Conference Series* (2021), vol. 1740, IOP Publishing, p. 012028.
- [31] Srebre, M., Schmolz, P., Hashemi, H., Ritter, M., and Kuhr, T. Generation of belle ii pixel detector background data with a gan. In *EPJ Web of Conferences* (2020), vol. 245, EDP Sciences, p. 02010.

# Appendix A: Ablation Study

We investigated the performance of our methods by removing both refraction and systematic corrections. Initially, raw  $\theta_c$  values without refraction correction produced lower  $\sigma_\beta$  compared with corrected ones (Fig. 9.1). The advantage was consequently lost upon the application of a systematic bias correction. Despite a better initial bias, the prediction variance was higher without refraction correction.

This also explains the decrease in quality after applying systematic correction to Hough baseline. As Hough baseline does not account for refraction, it demonstrates better compliance initially at the cost of a very high variance.

Next, we computed the mass distribution using Median values without the systematic correction.  $\pi/K$ -separation suffered tremendously, with the false positive and negative rates in the range of 50% and higher (Fig. 9.2, left).

The ablation study provided important evidence regarding the positive impact of our corrections in achieving the reported quality.

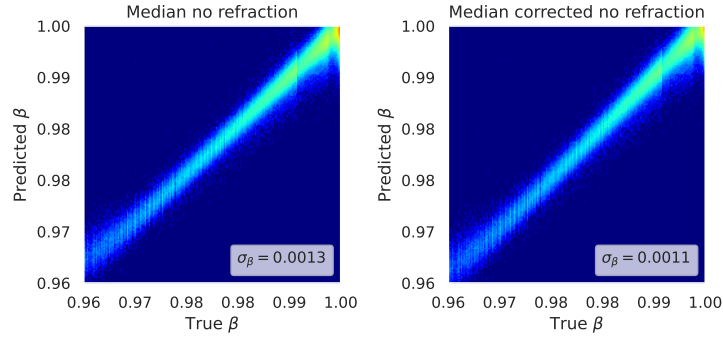


Figure 9.1: Balanced dataset 2d correlation for Median with and without systematic correction. Refraction correction was not performed.

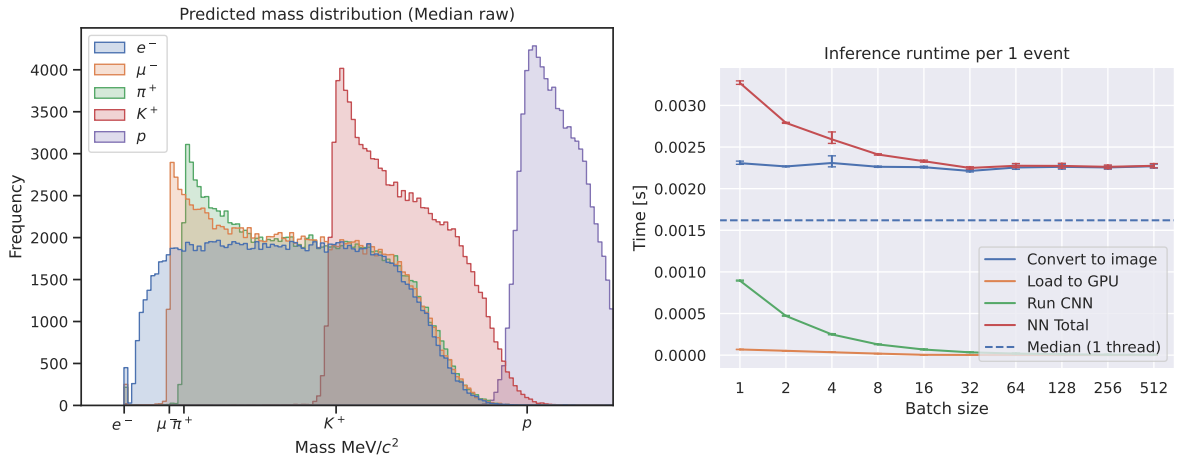


Figure 9.2: Left to right: predicted mass distribution with refraction, but without systematic correction, inference runtime per sample for the NN model.

Additionally, by far the most computationally expensive part of the pipeline was Cherenkov angle computation and coordinate processing prior to feeding an image into a neural network (Fig. 9.2, right). Extra Channel NN did not suffer from this as severely, nevertheless, it also used an expensive coordinate transformation in the Fourier features computation. We remedied this issue by implementing data parallelism, achieving a speed-up largely proportional to the number of computational nodes (up to 19x on 30 CPU cores). More efficient algorithms may be developed for a faster inference.

## Appendix B: Additional Figures

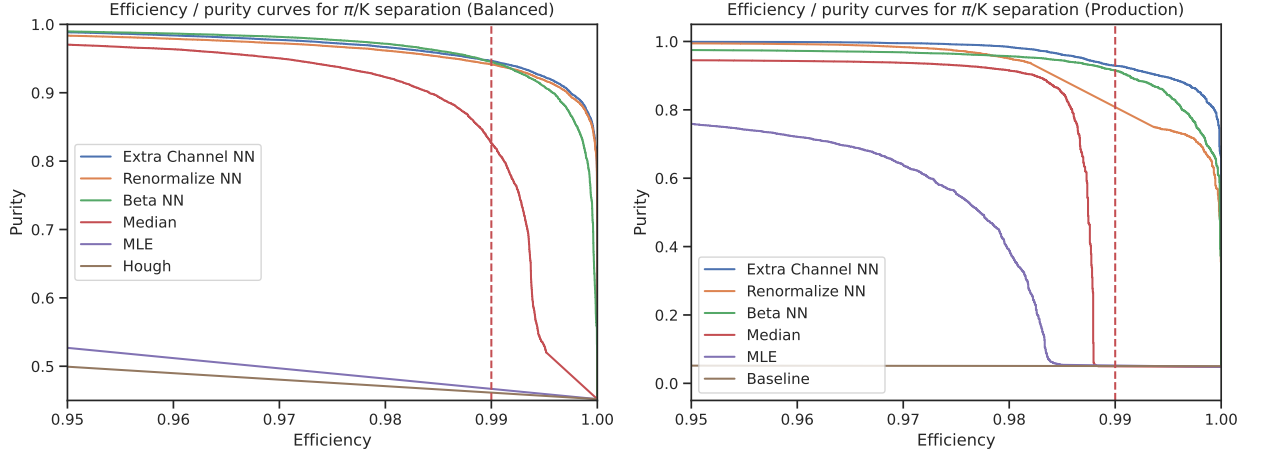


Figure 9.3:  $\pi/K$  efficiency / purity curves (precision-recall curves). The horizontal axis is zoomed in. The vertical line denotes

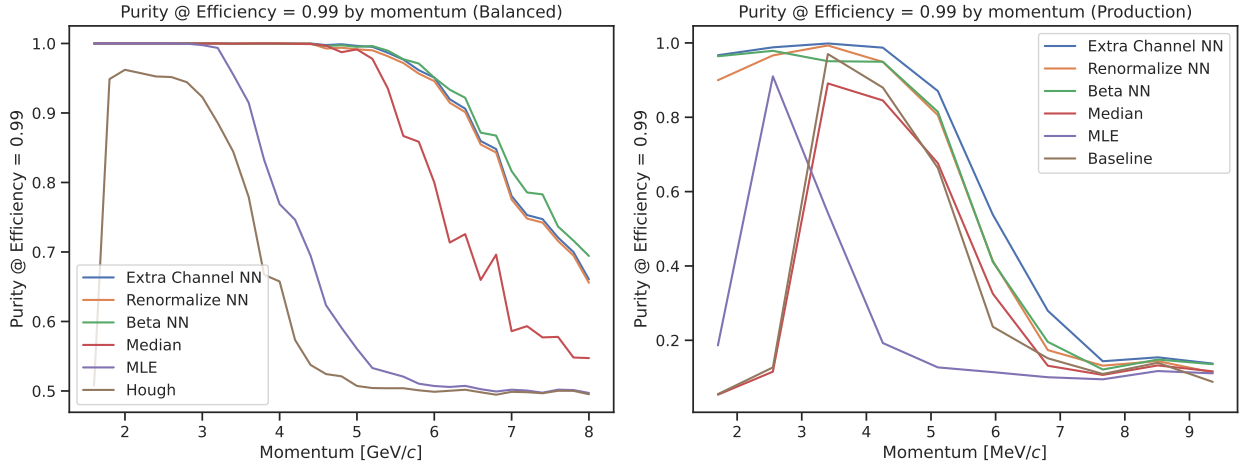


Figure 9.4:  $\pi/K$  classification quality binned by momentum. Left – Balanced, right – Production dataset.



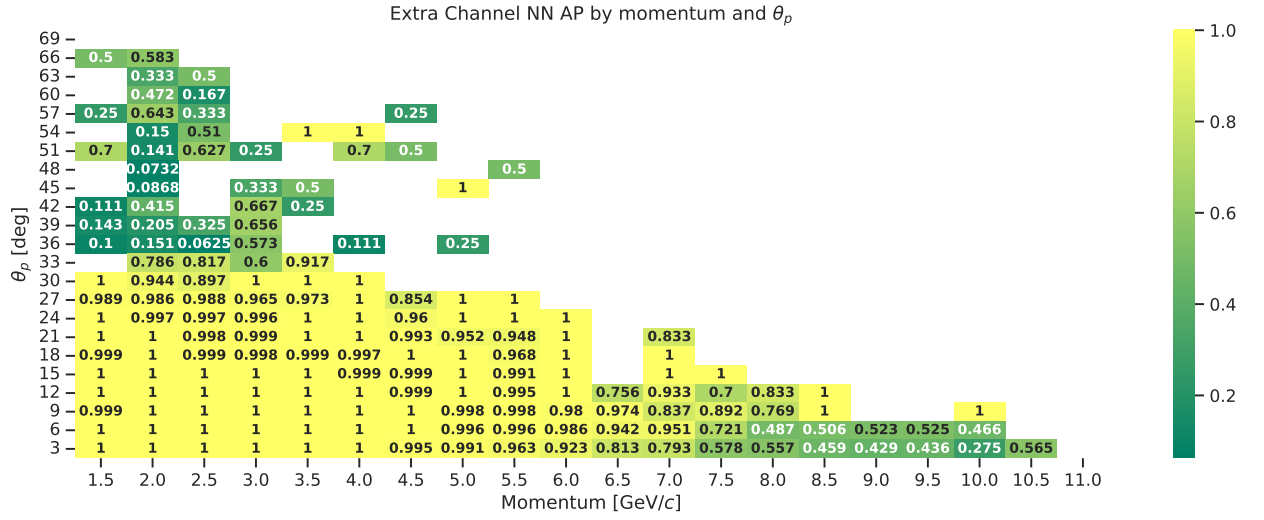


Figure 9.5: Detailed Production Extra Channel NN performance binned by momentum and  $\theta_p$ .

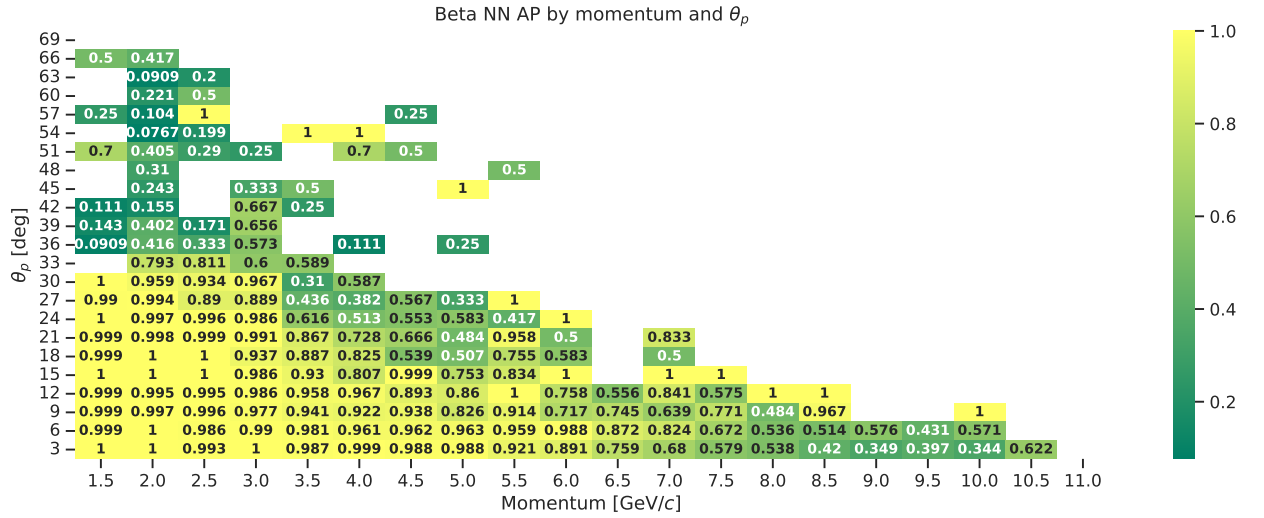


Figure 9.6: Detailed Production Beta NN performance binned by momentum and  $\theta_p$ .

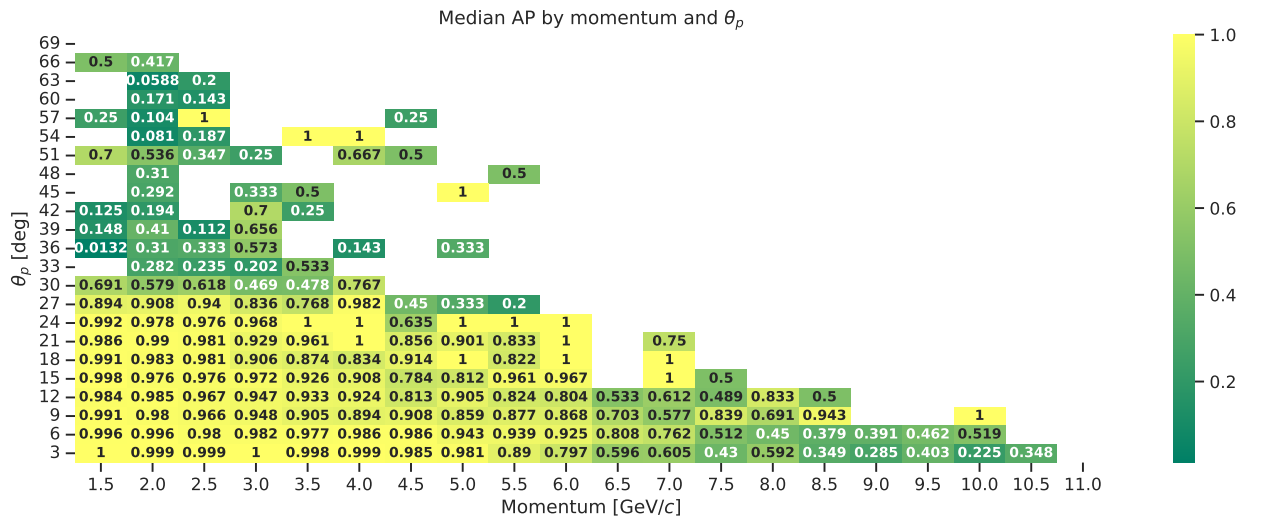


Figure 9.7: Detailed Production Median performance binned by momentum and  $\theta_p$ .