

Improved Particle Identification with the Belle II Calorimeter Using Machine Learning

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und
Naturwissenschaften Fachbereich Physik
der Universität Hamburg

vorgelegt von

Abtin Narimani Charan

aus Teheran

Hamburg

2023

Gutachter/innen der Dissertation:

Prof. Dr. Johannes Haller
Prof. Dr. Torben Ferber

Zusammensetzung der Prüfungskommission:

Prof. Dr. Johannes Haller
Prof. Dr. Torben Ferber
Prof. Dr. Jochen Liske
Prof. Dr. Gudrid Moortgat-Pick
Prof. Dr. Kerstin Tackmann

Vorsitzende/r der Prüfungskommission:

Prof. Dr. Jochen Liske

Datum der Disputation:

09.02.2024

Vorsitzender Fach-Promotionsausschusses PHYSIK:

Prof. Dr. Markus Drescher

Leiter des Fachbereichs PHYSIK:

Prof. Dr. Wolfgang J. Parak

Dekan der Fakultät MIN:

Prof. Dr.-Ing. Norbert Ritter

Abstract

This dissertation revolves around the utilization of Convolutional Neural Networks (CNNs) to advance Particle Identification (PID) within the Belle II Electromagnetic Calorimeter (ECL). The core goal of the research is to refine the differentiation process between low-momentum muons and charged pions. The ECL plays a significant role in the PID system as it is engineered to measure the energy deposition by both charged and neutral particles. The task of identifying low-momentum muons and charged pions within the ECL becomes particularly vital when they fail to reach the outer muon detector. In order to provide optimal data, the study employs track-seeded cluster energy images. The energy deposition patterns for muons and charged pions, as detected within crystals surrounding an extrapolated track at the ECL's entry point, are integrated with crystal positions in the $\theta - \phi$ plane along with the track's transverse momentum. This amalgamation of information is then utilized to train the CNN, capitalizing on the distinctiveness between the dispersed energy depositions of pion hadronic interactions and the more localized muon electromagnetic interactions. The study includes a comparison of the CNN algorithm's performance with other PID methods currently in use at Belle II, which predominantly rely on track-matched clustering information. The findings imply that the CNN PID method improves the separation between muons and charged pions in low-momentum regions. The research includes samples with varying beam backgrounds, including no beam background. The effectiveness of the CNN method has been assessed with different energy thresholds for ECL crystals, utilizing 21.5 fb^{-1} data from 2020 and 2021 and Monte Carlo (MC) samples. To substantiate the CNN method with real data, clean samples of muons and charged pions have been singled out using $e^+e^- \rightarrow \mu^+\mu^-\gamma$ and $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$, respectively. Finally, recognizing that the CNN is sensitive to tracks in close proximity within a single event, additional research was conducted to evaluate the CNN's performance with isolated and non-isolated tracks within the ECL.

Zusammenfassung

Diese Dissertation befasst sich mit der Nutzung von Convolutional Neural Networks (CNNs) zur Verbesserung der Teilchenidentifizierung (PID) im Belle II elektromagnetischen Kalorimeter (ECL). Das Hauptziel der Forschung ist es, den Differenzierungsprozess zwischen langsamen Myonen und geladenen Pionen zu verfeinern. Das ECL spielt eine bedeutende Rolle im PID-System, da es darauf ausgelegt ist, die Energieablagerung sowohl von geladenen als auch von neutralen Teilchen zu messen. Die Aufgabe, langsame Myonen und geladene Pionen innerhalb des ECL zu identifizieren, wird besonders wichtig, wenn diese den äußeren Myondetektor nicht erreichen. Um optimale Daten zu liefern, verwendet die Studie durch Spuren gesäte Cluster-Energiebilder. Die Energieablagerungsmuster für Myonen und geladene Pionen, die in den Kristallen um eine extrapolierte Spur am Eingangspunkt des ECLs detektiert werden, werden mit den Kristallpositionen in der $\theta - \phi$ -Ebene zusammen mit dem transversalen Impuls der Spur integriert. Diese Kombination von Informationen wird dann genutzt, um das CNN zu trainieren, wobei die Unterschiedlichkeit zwischen den verteilten Energieablagerungen der hadronischen Wechselwirkungen des Pions und den lokalisierteren elektromagnetischen Wechselwirkungen des Myons ausgenutzt wird. Die Studie beinhaltet einen Vergleich der Leistung des CNN-Algorithmus mit anderen derzeit bei Belle II verwendeten PID-Methoden, die überwiegend auf spurgepaarten Clustering-Informationen basieren. Die Ergebnisse deuten darauf hin, dass die CNN-PID-Methode die Trennung zwischen Myonen und geladenen Pionen in Bereichen niedriger Impulse verbessert. Die Forschung umfasst Proben mit verschiedenen Strahlhintergründen, einschließlich keines Strahlhintergrunds. Die Wirksamkeit der CNN-Methode wurde mit verschiedenen Energieschwellen für ECL-Kristalle bewertet, wobei $21,5 \text{ fb}^{-1}$ Daten von 2020 und 2021 und Monte-Carlo (MC)-Proben verwendet wurden. Um die CNN-Methode mit echten Daten zu untermauern, wurden saubere Proben von Myonen und geladenen Pionen unter Verwendung von $e^+e^- \rightarrow \mu^+\mu^-\gamma$ und $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$ isoliert. Schließlich, in der Erkenntnis, dass das CNN empfindlich auf Spuren in unmittelbarer Nähe innerhalb eines einzelnen Ereignisses reagiert, wurde zusätzliche Forschung durchgeführt, um die Leistung des CNNs bei isolierten und nicht isolierten Spuren innerhalb des ECL zu bewerten.

“Science cannot solve the ultimate mystery of nature. And that is because, in the last analysis, we ourselves are part of nature and therefore part of the mystery that we are trying to solve.”

Max Planck

Table of Contents

1 Introduction	1
2 Theory and Physics Motivation	3
2.1 The Standard Model	4
2.2 Significance of Separating Muons and Charged Pions	6
2.2.1 Semi-leptonic Decays of B Mesons	7
2.2.2 Measurements of the CKM Matrix Elements V_{ub}, V_{cb}	7
2.2.3 Cross Section Measurements of $e^+e^- \rightarrow \pi^+\pi^-$	8
2.3 Particle Interactions in Calorimeters	9
3 The Belle II Experiment	15
3.1 SuperKEKB	15
3.2 The Belle II Detector	17
3.3 Belle II Analysis Software Framework	21
3.4 Electromagnetic Calorimeter	23
3.5 Charged Particle Identification at Belle II	26
4 Convolutional Neural Networks	31
4.1 Convolution	32
4.2 Fully Connected Layers	33
4.3 Embedding	34
4.4 Activation Function	35
4.5 Preprocessing	35
4.6 Datasets: Training, Validation, and Test	36
4.7 Backpropagation	36
4.8 Hyperparameters	38
4.9 Regularization	39

5 Particle Identification with the CNN	41
5.1 Event Generation and Reconstruction	41
5.2 Event Selection	42
5.3 Inputs and Preprocessing	43
5.4 Training the CNN	48
5.4.1 Hyperparameter Optimization	48
5.4.2 Architecture and Training	51
5.5 Model Evaluation	53
5.5.1 Muon Efficiency vs. Pion Fake Rate	55
5.5.2 Evaluation in Different Transverse Momentum Ranges	57
5.6 Comparison with Other PID Methods	61
5.6.1 Efficiencies and Fake Rates at Fixed Values	61
5.6.2 Evaluation in Different Transverse Momentum Regions	62
5.6.3 Comparative Analysis of CNN and BDT in the ECL Barrel	64
5.7 Energy Thresholds' Effect	67
5.7.1 Significance of Energy Thresholds	67
5.7.2 Trade-off Between Efficiency and Fake Rate	68
5.7.3 Performance Across Transverse Momentum Ranges	70
5.7.4 Implications on Event Complexity	70
5.8 Beam Backgrounds' Effect	73
5.8.1 Testing on Different Beam Backgrounds	73
5.8.2 Training on Different Beam Backgrounds	75
6 CNN Validation with Data	81
6.1 Collision Dataset	81
6.2 MC Simulated Samples	81
6.3 $e^+e^- \rightarrow \mu^+\mu^-\gamma$	82
6.3.1 Preselection Criteria	82
6.3.2 Event Selection	83
6.3.3 Data and MC Comparison	83
6.4 $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$	88
6.4.1 Preselection Criteria	88
6.4.2 Event Selection	88

6.4.3 Data and MC Comparison	88
6.5 Comparison of CNN Outputs	95
6.6 Comparison of CNN and Other PID Methods	98
6.6.1 ROC Curve Analysis	98
6.6.2 Efficiency and Fake Rate Comparison	98
6.7 CNN Evaluation on Energy Thresholds	105
6.8 CNN Evaluation with Various Beam Backgrounds	107
6.9 Isolation Effects	109
6.9.1 Mis-classification Analysis of Muon Tracks	110
6.9.2 Evaluating the Impact of Isolation Cuts	111
6.9.3 Improving CNN Performance with Isolation Cuts	111
7 Summary and Outlook	115
List of Figures	119
List of Tables	123
Acronyms	125
Bibliography	129
Disclaimer	137
Acknowledgements	139

Chapter 1

Introduction

Particle physics has always been at the cutting edge of scientific discovery, making astounding progress in our understanding of the most fundamental building blocks of matter and their complex interactions. It's the Standard Model (SM) of particle physics that has been our guiding star - a powerhouse of a theory that brings together the electromagnetic, weak, and strong interactions. It has proven to be incredibly successful, predicting an impressive range of phenomena we've observed. However, as extraordinary as the SM is, it's not perfect. It struggles when it comes to incorporating gravity and falls short in explaining some of the universe's most mysterious phenomena, like dark matter or the fact that matter seems to far outweigh antimatter everywhere we look. These are the big, burning questions that remain unanswered, and they are what drive us to venture beyond the SM in search of new physics.

One of the primary tools employed in this pursuit is the particle detector, an ingenious device instrumental in tracking and identifying particles generated from high-energy collisions. The Belle II experiment, situated at the SuperKEKB accelerator in Tsukuba, Japan, is a prominent scientific endeavor specifically designed to explore new frontiers in particle physics. The Belle II detector is an intricate apparatus comprised of multiple components, each dedicated to measuring distinct properties of particles. Among these components, the Electromagnetic Calorimeter (ECL) stands out, effectively gauging particle energy by absorbing their emissions and generating detectable signals [1].

Particle Identification (PID) assumes paramount significance in the analysis of particle physics data. Accurate identification of particles allows physicists to meticulously reconstruct the processes occurring within particle collisions. However, PID poses a formidable challenge, particularly for low-momentum particles that do not reach the outer part of the detector. This challenge is particularly relevant in the context of the Belle II ECL, where low-momentum muons and charged pions are the focal point of this dissertation [1].

In recent years, Machine Learning (ML) techniques, and in particular, Convolutional Neural Networks (CNNs), have emerged as powerful tools in advancing PID capabilities within the particle detectors. CNNs, a subtype of deep learning algorithms, exhibit exceptional ability in tasks such as image and pattern recognition. Within the ECL domain,

CNNs can be trained to discern distinctive energy deposition patterns associated with different particles, thereby elevating the precision of PID [2, 3].

The primary objective of this dissertation is to employ track-seeded cluster energy images, which provide optimal data for CNN analysis. The performance of the CNN-based PID method will be evaluated and compared against other PID methods utilized at Belle II, which rely on track-matched clustering information [2, 3].

The structure of this dissertation is organized as follows:

- Chapter 2 provides an overview of the SM, relevant physics channels benefiting from improved μ/π separation in the context of Belle II. Then calorimeters and particle interactions when passing through detector will be explained.
- Chapter 3 introduces the Belle II detector with a brief introduction about its sub-detectors, focusing on the ECL. It explores the detector's structure, the Belle II software, and PID methods that are currently used.
- Chapter 4 explains concepts related to CNNs that are pertinent to this dissertation.
- Chapter 5 constitutes the core of this dissertation, fully describing the methodology of using CNNs for μ/π separation in the ECL. The chapter begins by detailing event generation and selection, followed by listing inputs and preprocessing steps. It proceeds to explain the training and testing of the CNN method, comparing it with other PID methods in the Belle II software. The chapter also examines the effects of varying thresholds and beam backgrounds.
- Chapter 6 focuses on validating the CNN approach through the examination of charged pion and muon samples derived from D^* decay and radiative dimuon events, respectively. It provides a detailed account of the selection process for charged pions and muons in both Monte Carlo (MC) simulations and collision data gathered during 2020 and 2021. This chapter includes a comparative analysis of various PID methods as applied to the aforementioned data and MC. It also delves into how varying energy thresholds and different beam background conditions impact the efficacy of the CNN. The chapter concludes by assessing the influence of isolated and non-isolated tracks within the ECL on the performance of the CNN.
- Chapter 7 provides a summary, draws conclusions, and outlines potential avenues for future studies.

Chapter 2

Theory and Physics Motivation

A Tribute to the Unforgettable

The first decades of the 20th century were instrumental in shaping the trajectory of what we now understand as modern physics. Breakthroughs such as the advent of quantum theory and the special theory of relativity laid the foundation for our contemporary comprehension of the universe. The journey commenced in 1901 with Max Planck introducing the concept of energy quantization as a solution to the ultraviolet catastrophe, an inherent paradox in classical physics related to black body radiation [4]. Planck proposed that energy is emitted or absorbed in discrete packets, termed as quanta.

Then, in 1905, Albert Einstein introduced his special theory of relativity [5], a vital component for the development of quantum theory. Subsequently, in 1913, Niels Bohr proposed a model of the atom wherein electrons occupy quantized orbits around the nucleus [6], thereby contributing another foundational element to quantum theory.

In 1918, Emmy Nöther proposed a revolutionary concept known as Nöther's theorem. Nöther's theorem asserts that each continuous symmetry indicates the presence of a conserved quantity [7]. This key insight holds monumental significance in the subsequent progression and maturation of quantum mechanics.

The decade of the 1920s witnessed crucial advancements by Louis de Broglie [8], Werner Heisenberg [9], Erwin Schrödinger [10], Paul Dirac [11], Max Born [12], and Wolfgang Pauli [13]. Their efforts led to a more comprehensive formulation of quantum mechanics. By the 1930s, Enrico Fermi added to the mix, introducing the first theory about the weak force in atoms, explaining why some atoms break apart [14].

During the first half of the 20th century, discoveries of various particles spurred the development of Quantum Field Theories (QFTs) which combined quantum mechanics and special relativity to explain the properties and interactions of subatomic particles. Quantum Electrodynamics (QED), the QFT that describes electromagnetism, was successfully formulated by Richard Feynman [15], Julian Schwinger [16], and Sin-Itiro Tomonaga [17].

By the mid-20th century, the accumulated knowledge set the stage for the revelation

of the SM. Key to its development was Murray Gell-Mann, who in the 1960s introduced quarks, elementary particles that form the building blocks of matter [18].

Today, we owe much of our understanding of the universe to these theoretical and experimental physicists.

2.1 The Standard Model

The SM presents a comprehensive catalogue of fundamental particles and outlines the interaction mechanisms among them. It effectively integrates the principles of quantum mechanics and special relativity to construct a well-grounded QFT. In the chronological context of the 1960s, three distinguished physicists—Steven Weinberg, Abdus Salam, and Sheldon Glashow—individually put forth a theoretical proposition to combine two fundamental interactions, specifically the electromagnetic and weak interactions [19, 20, 21]. This proposed unification, now formally referred to as the electroweak interaction, has substantially augmented our comprehension of particle physics. The unprecedented significance of their work was acknowledged globally when they were conferred with the Nobel Prize in Physics in 1979 [22].

The SM essentially revolves around two core ideas: particles and forces. Particles are the raw materials that build visible matter, while forces dictate how these particles interact. The model classifies particles into fermions and bosons. It covers the electromagnetic, weak, and strong interactions, but gravity, the fourth elemental force, sits outside its realm. Despite its profound insights, the SM has its limitations, with mysteries like dark matter, dark energy, and the matter-antimatter imbalance yet to be unraveled [22].

Fermions

Fermions, with non-integer spin, adhere to the Pauli exclusion principle [13]. They split into quarks and leptons. Quarks come in six flavors: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). They possess fractional electric charges and participate in strong and weak interactions. Due to confinement, quarks are never isolated but form composite particles called hadrons [22].

Leptons, on the other hand, do not engage in strong interactions. They comprise the electron (e), muon (μ), tau (τ), and their respective neutrinos (ν_e, ν_μ, ν_τ). Each charged lepton comes with a corresponding neutrino, which interacts solely through the weak force. While the electron, muon, and tau have charges, neutrinos are neutral [22].

Bosons

Bosons, with integer spins, mediate the fundamental forces. The SM identifies five bosons: the photon (γ), W and Z bosons, gluons (g), and the Higgs boson (H). The photon mediates electromagnetic interactions among charged particles, while the W and Z bosons

mediate the weak interaction, crucial for radioactive decay and nuclear fusion. Gluons bind quarks within hadrons [22].

The Higgs boson, a distinct particle discovered in 2012 by the ATLAS and CMS experiments at CERN’s Large Hadron Collider (LHC) [23, 24], is responsible for the generation of the Higgs field, a scalar field spreading through all of space. It has spin 0 and is electrically neutral. It is measured to have a mass of $125.10 \pm 0.14 \text{ GeV}/c^2$. Other fundamental particles acquire mass by interacting with Higgs field, an effect termed the Higgs mechanism [25, 26, 27].

Figure 2.1 presents a detailed schematic diagram of the elementary particles according to the SM. This illustration includes comprehensive information on each particle, such as mass, charge, and spin, offering a clear and complete representation of their fundamental properties.

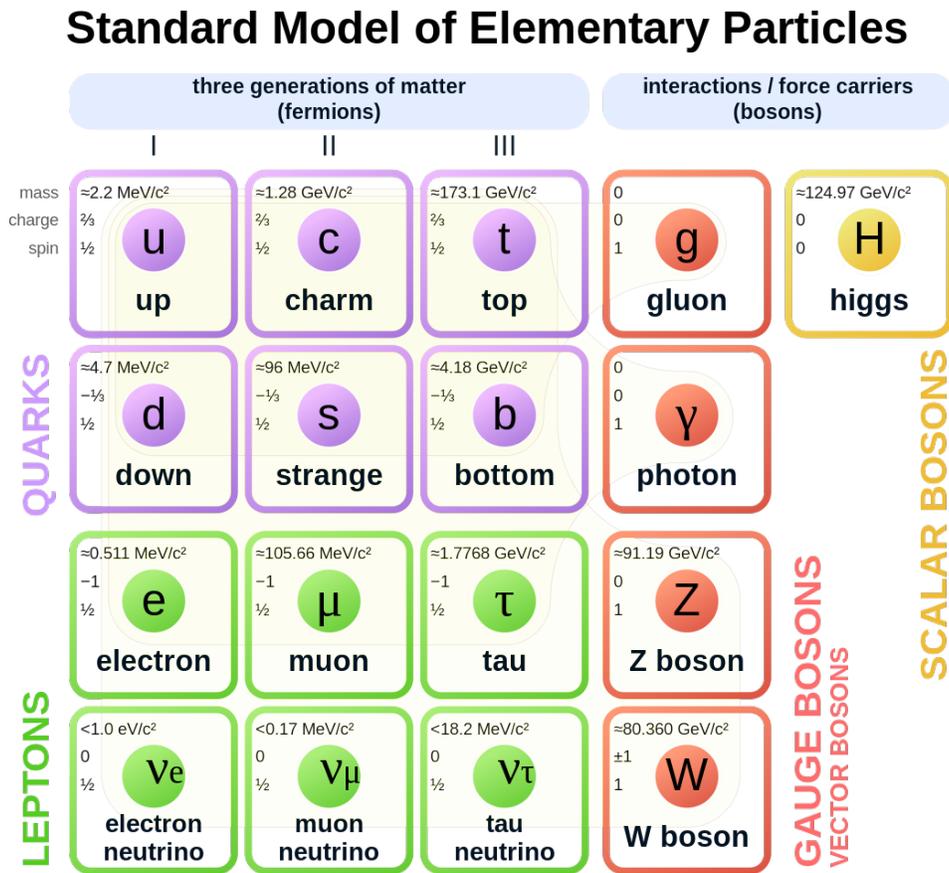


Figure 2.1: The SM of elementary particle including matter fields or fermions (violet and green squares) and bosons (red and yellow squares). Mass, charge, and spin of the particle are written at the upper left part of the squares. Image from [28].

Strong Interaction

The strong interaction is the primary focus of Quantum Chromodynamics (QCD). This force is responsible for binding quarks together within hadrons, such as protons and neutrons. Gluons, which carry the color charge, mediate the strong interaction. Notably, this force is powerful enough to overcome the electromagnetic repulsion between positively charged protons within atomic nuclei, ensuring nuclear stability [29].

Weak Interaction

The weak interaction plays a critical role in processes like beta decay ($p^+ \rightarrow n^0 + e^+ + \nu_e$). It is mediated by the W and Z bosons. Unique in its ability to change quark flavors, the weak interaction enables transformations between different types of particles [29]. Its integration with electromagnetism into the electroweak theory represents a pivotal development in particle physics [19, 20, 21].

Electromagnetic Interaction

The electromagnetic force, a cornerstone of classical and quantum physics, is mediated by photons (γ). This interaction operates between particles that possess electric charge. Its scope ranges from influencing the structure of atoms to governing the behavior of light. QED offers the most precise description of this force [30].

2.2 Significance of Separating Muons and Charged Pions

Within collider experiments, accurate identification of particles plays a crucial role in deciphering the underlying processes and interactions. Among the various challenges faced by experimentalists, the separation of muons and charged pions stands out as a particularly significant one. Both muons and charged pions are frequently produced in high-energy collisions, and their behaviours through detectors can sometimes appear deceptively similar. However, their distinct roles in various physical processes necessitate their clear differentiation.

The μ/π separation is crucial for the following measurements, important for this dissertation:

- Background noise reduction in semi-leptonic decays of B mesons,
- Measurements of the CKM matrix elements V_{ub} , V_{cb} ,
- Cross section measurements of $e^+e^- \rightarrow \pi^+\pi^-$.

2.2.1 Semi-leptonic Decays of B Mesons

The study of semi-leptonic decays of B mesons, such as $b \rightarrow c\tau^+\nu_\tau$ transitions and those involving $b \rightarrow s\mu^+\mu^-$ and $b \rightarrow se^+e^-$, plays a crucial role in exploring potential Lepton Flavour Universality (LFU) violations and uncovering new physics beyond the SM. In the Belle II experiment, focus is given to the reconstruction of tau leptons from their decay products, notably $\tau^- \rightarrow \mu^-\bar{\nu}_\mu\nu_\tau$ or $\tau^- \rightarrow e^-\bar{\nu}_e\nu_\tau$. The momentum spectrum of these resulting light leptons, both electrons and muons, is predominantly low [3].

Challenges arise in PID within this low momentum range. The primary method used is analyzing the E/p ratio, where E is the energy measured in the ECL and p is the reconstructed momentum of the charged track. This ratio, while effective for particles with $p > 1$ GeV/ c , becomes less reliable for lower momentum electrons due to increased energy losses from bremsstrahlung. Moreover, muon identification in the outermost detector faces lower efficiency for low momentum muons, often falling outside the acceptance of the detectors. Other sub-detectors, such as the time of propagation detector (3.2) and the aerogel ring-imaging Čerenkov detector (3.2), also struggle with efficient μ/π separation in this momentum range, due to their masses being near identical [1].

Therefore, precise separation of muons and charged pions at these lower momenta is not just a technical challenge but also a crucial element in accurately interpreting semi-leptonic B decay data. This separation is vital for reducing background noise from hadronic processes and enhancing the clarity of leptonic signals, playing a pivotal role in investigating LFU violations.

2.2.2 Measurements of the CKM Matrix Elements V_{ub} , V_{cb}

Particle physics delves deep into the study of fundamental symmetries and their occasional violations. Of these symmetries, Charge conjugation (C), parity (P), and their combined form (CP) have attracted significant attention. The C symmetry posits that the laws of physics remain consistent when a particle is swapped with its antiparticle. Similarly, the P symmetry maintains that these laws are invariant when a system undergoes a spatial mirror reflection. When combined, CP symmetry suggests that physics should remain unchanged under both transformations simultaneously. However, this is not the case in weak interactions, leading to the phenomenon known as CP violation [29, 31].

The 1960s witnessed a groundbreaking discovery of CP violation in the decays of neutral K mesons [32]. As a consequence of this revelation, the Cabibbo–Kobayashi–Maskawa (CKM) theory was formulated. This theory posits CP violation as an intrinsic feature of the SM. According to the CKM theory, CP violation arises from a complex phase in the quark mixing matrix, a hypothesis that can be tested in B meson experiments [33].

The CKM matrix elucidates the mixing between the three generations of quarks. It plays an important role in understanding the weak interactions between quarks. Within this matrix, the elements V_{ub} and V_{cb} stand out due to their significance in testing the CKM sector of the SM and complementing the measurements of CP asymmetries in B decays [34].

The CKM matrix is represented as:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}. \quad (2.1)$$

The complex phase in this matrix, typically parameterized by the Wolfenstein parameters, is the source of CP violation in the SM. The parameter η associated with the imaginary part is responsible for this CP violation [22, 35].

Diving deeper into the CKM matrix, the semileptonic decays $b \rightarrow u(c)l\nu$ offer insights into the complex realm of quark transitions. Here, a b quark transitions to a u or c quark, simultaneously emitting a lepton and a neutrino. The momentum of the emitted lepton is key to understanding the quark transition. For precise measurements of V_{ub} and V_{cb} , it's essential to accurately identify this lepton, especially at low momentum. Furthermore, the branching ratios of these decays, representing the fraction of B mesons decaying through these channels, are closely tied to the values of V_{ub} and V_{cb} . Hence, accurate lepton identification is paramount for the correct measurement of these branching ratios [1, 34].

2.2.3 Cross Section Measurements of $e^+e^- \rightarrow \pi^+\pi^-$

The cross section of the process $e^+e^- \rightarrow \pi^+\pi^-$ is a crucial area of research in particle physics, with significant contributions from experiments such as KLOE [36, 37, 38], BABAR [39, 40], BESIII [41], CLEO-c [42], CMD-2 [43], and SND [44]. These measurements are crucial for delving into low-energy QCD dynamics and are essential for probing phenomena like chiral perturbation theory and the dynamics of resonances such as the ρ meson.

To accurately measure the $e^+e^- \rightarrow \pi^+\pi^-$ cross section, it is essential to study a scan of the $\pi\pi$ -invariant-mass spectrum. This can be achieved using two primary methods. The first is the scan method, employed by CMD-2 and SND, where the Center-of-Mass (CM) energy of the collision is directly adjusted to various energies. While this method offers precise control over the collision energy, it is limited by the need for high beam energy resolution and results in discrete data points with gaps in the spectrum. The second method involves using Initial State Radiation (ISR), also known as radiative return. This approach, utilized by KLOE, BABAR, BESIII, and CLEO-c, involves an electron or positron emitting a photon before annihilation, effectively reducing the collision energy and allowing a continuous scan of a wide energy range at a fixed beam energy [37, 39, 41, 42].

The most precise measurements of the $e^+e^- \rightarrow \pi^+\pi^-$ cross section have been conducted by KLOE and BABAR. However, discrepancies between these measurements have led to unresolved tensions and underscored the need for new, precise measurements of this cross section [37, 39].

The Belle II experiment emerges as a promising candidate to resolve these discrepancies and enhance the precision of the Hadronic Vacuum Polarization (HVP) calculations. HVP is a major component in the theoretical prediction of the muon’s anomalous magnetic moment $g - 2$. Precise measurements of the $e^+e^- \rightarrow \pi^+\pi^-$ cross section contribute significantly to reducing the uncertainty in HVP calculations. Similarly, Hadronic Light-by-Light scattering (HLbL) contributions, another critical aspect of the muon $g - 2$ calculations, are influenced by the accuracy of these cross section measurements. The HLbL is a subtle yet important phenomenon in QFT and its precise understanding hinges on accurate experimental data [45].

In Belle II, the ability to differentiate muons from charged pions at low momentum is crucial, especially given that a significant portion of the two pion mass spectrum lies in the low energy region ($< 1 \text{ GeV}/c^2$). This region is particularly sensitive to new physics phenomena and contributes substantially to the HVP and HLbL calculations. Improved particle identification at low momentum can lead to more accurate cross section measurements, aiding in the reduction of uncertainties in HVP and HLbL calculations, refining our understanding of these phenomena, and potentially unveiling signs of new physics beyond the SM [1].

2.3 Particle Interactions in Calorimeters

Particle interactions in calorimeters are pivotal in high-energy physics research. Calorimeters are tailored to measure particle energy based on the principle of energy conservation. A cornerstone for understanding these interactions is the Bethe-Bloch equation, which elucidates the energy loss of charged particles as they traverse different materials [46].

Formulated by Hans Bethe and Felix Bloch in the early 20th century [47, 48, 49], the Bethe-Bloch equation quantifies the average energy loss rate of a charged particle moving through matter. The mean rate energy loss by moderately relativistic charged heavy particle is described as [50]:

$$\left\langle -\frac{dE}{dx} \right\rangle = K z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \left(\frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} \right) - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right], \quad (2.2)$$

where dE/dx represents the energy loss per unit distance traveled by the particle, K is a constant equals to $4\pi N_A r_e^2 m_e c^2$ ($0.307075 \text{ MeV mol}^{-1} \text{ cm}^2$), z is the charge number of incident particle, Z is the atomic number of absorber, A is the atomic mass of the absorber, β is the velocity of the particle in units of the speed of light, γ is the Lorentz factor, I is the mean excitation energy, $\delta(\beta\gamma)$ represents density effect correction to ionization energy loss, and W_{\max} is the maximum kinetic energy transfer to the atomic electrons in a single collision [50]. For a particle with mass M ,

$$W_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2m_e/M + (m_e/M)^2}. \quad (2.3)$$

At lower momenta, particles with a larger mass demonstrate a pronounced dE/dx , a behavior that aligns with the $1/\beta^2$ component in Bethe-Bloch equation. This trend is consistent with the general understanding derived from the Bethe-Bloch equation. An intriguing point on the Bethe-Bloch curve is the value of the kinetic energy corresponding to the minimal energy loss. Particles traveling at velocities corresponding to this minimum are termed Minimum Ionizing Particles (MIPs). Muons and charged pions, given their intermediate masses between electrons and heavier particles like protons, often behave as MIPs under certain conditions. Specifically, muons, due to their relatively higher mass compared to electrons, tend to be minimum ionizing over a broad range of energies. This makes them particularly significant in many experimental setups, as their consistent energy loss rate simplifies detection and measurement. Charged pions, on the other hand, have a slightly different behavior but can still be considered MIPs in specific energy ranges. As depicted in Figure 2.2, when a particle's velocity decreases, its energy loss due to ionization surges dramatically. This phenomenon is especially evident for heavier particles like protons, deuterons, and α particles, which display a notable peak in energy loss as they approach rest. As a particle undergoes ionization and nears this region of sharp energy loss, a self-amplifying cycle is initiated. This results in the particle rapidly depleting its remaining energy within a confined distance, a crucial aspect to consider in applications like calorimetry in high-energy physics experiments [35, 46, 50].

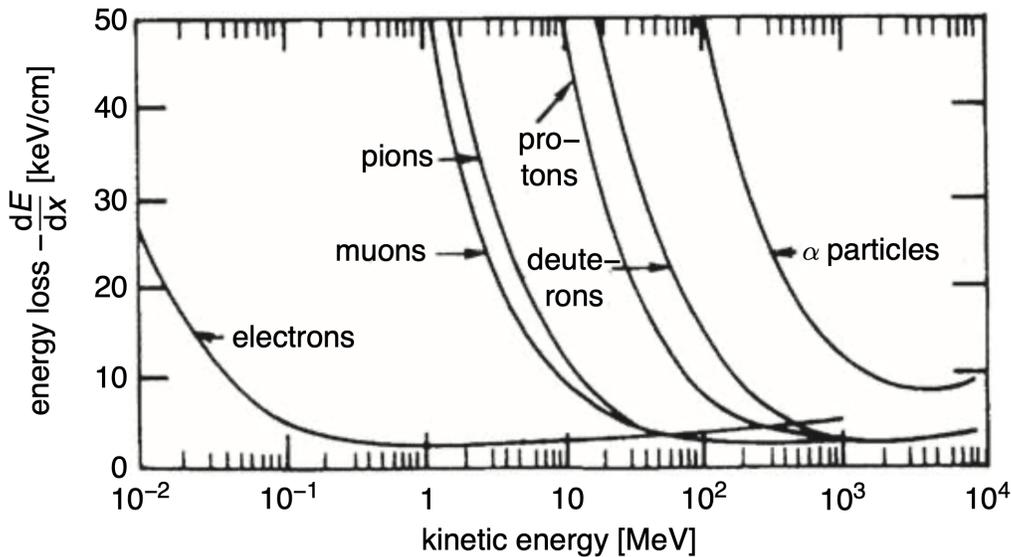


Figure 2.2: Ionization energy loss for various particles in air as a function of momentum. The curve shows minimum energy loss for certain particle velocities, corresponding to MIPs. Image from [46].

The Bethe-Bloch equation allows us to understand the energy loss of charged particles due to their interactions with the atomic electrons of the material they traverse. This equation is of significant importance in calorimetry, as it provides insights into the energy deposition of particles within the calorimeter material. By knowing the energy loss of particles, one can accurately determine their energy and track their trajectories within the

detector [51, 52]. In high-energy physics experiments, calorimeters aim to stop particles entirely, ensuring that they deposit all their energy within the detector. This is achieved by carefully selecting materials with suitable properties and optimizing the thickness of the calorimeter. The Bethe-Bloch equation helps in the design and optimization process by providing information about the energy loss of particles in different materials and at various energies. Additionally, the equation aids in the identification of different particle types based on their energy loss patterns, contributing to PID algorithms and the study of specific decay channels [50].

By combining the principles of calorimetry, such as the formation of electromagnetic and hadronic showers, with the knowledge gained from the Bethe-Bloch equation, one can construct efficient calorimeters for precise energy measurements and PID in high-energy physics experiments [51, 52].

Calorimeters

Calorimeters are designed to stop particles, forcing them to deposit all their energy within the detector. This is achieved by causing the particles to undergo a series of interactions that result in the production of secondary particles, which in turn interact and produce more particles. This cascade of particles, known as a shower, continues until the energies of the individual particles are too low to produce further interactions. The energy deposited in the calorimeter is then collected and measured. This can be done directly, by measuring the heat produced by the particle shower, or indirectly, by detecting the light produced by the secondary particles as they pass through a suitable medium. There are two main types of calorimeters used in high-energy physics: electromagnetic calorimeters and hadronic calorimeters. Electromagnetic calorimeters are designed to measure the energy of electrons and photons, which primarily interact via the electromagnetic force. Hadronic calorimeters, on the other hand, are designed to measure the energy of hadrons, which interact via the strong nuclear force [46, 51, 52, 53].

Interactions in Calorimeters

The specific interactions that occur in calorimeters depend on the type of particle and the type of calorimeter. In electromagnetic calorimeters, the primary interactions are pair production and bremsstrahlung. In hadronic calorimeters, the primary interactions are nuclear reactions and particle decay [51].

Electron Interactions and Electromagnetic Showers

Pair production is the process by which a photon, upon interacting with a nucleus, produces an e^+e^- pair. The energy of the photon is converted into the mass and kinetic energy of the pair, with any excess energy carried away by the pair as kinetic energy. This process is a major contributor to the energy deposition in electromagnetic calorimeters.

Bremsstrahlung, on the other hand, is the process by which an electron, upon interacting with a nucleus, emits a photon. The energy of the electron is partially converted into the energy of the photon, with the electron not stopped. This process is responsible for the production of secondary photons in electromagnetic calorimeters, which can then undergo pair production and contribute to the electromagnetic shower [46, 50, 51].

Electromagnetic showers are initiated by high-energy electrons, positrons, or photons interacting with the calorimeter’s dense material. Electrons primarily lose energy through bremsstrahlung, emitting photons within a nucleus’s Coulomb field. These photons can then induce e^+e^- pair production, creating a cascade of secondary particles. The development of such showers is characterized by the radiation length (X_0), a characteristic distance over which a high-energy electron loses all but $1/e$ ($\approx 37\%$) of its energy by bremsstrahlung, and a high-energy photon has a $1/e$ chance of undergoing pair production. The showers proceed until the energy of the particles falls below critical thresholds for pair production and energy loss [22, 46]. A schematic showing the development of an electromagnetic shower is illustrated in Figure 2.3.

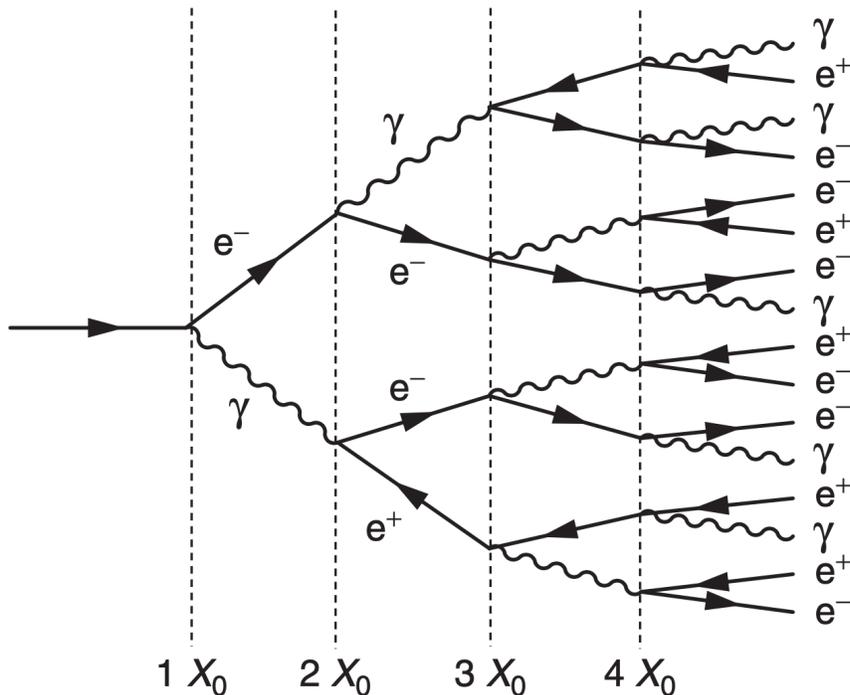


Figure 2.3: Progressive stages of an electromagnetic shower. The shower is initiated by a high-energy photon or electron interacting with matter. As depicted, the number of secondary particles (primarily electrons, positrons, and photons) produced in the shower roughly doubles after traversing each radiation length. Image from [22].

Hadronic Showers

Hadronic showers are more complex, starting with a hadron's strong interaction with the nuclei in the detector material. This leads to the production of secondary particles like pions, kaons, and neutrons, and can induce nuclear reactions such as fission or evaporation. Charged secondary particles further ionize the material and may cause additional nuclear interactions, forming a hadronic cascade. Energy loss in hadronic showers is classified into electromagnetic and hadronic components, with the latter comprising energy deposited by ionization from charged hadrons [22, 46].

Chapter 3

The Belle II Experiment

The Belle II experiment is a B -Factory located at the SuperKEKB e^+e^- collider in Tsukuba, Japan. It stands as a critical venture towards exploring CP violation with unrivaled precision, particularly in B meson decays. This experiment tests CKM theory predictions and potentially unveils new sources of CP violation beyond the SM, which is a crucial aspect to understand why there is more matter than antimatter in the Universe. While the Belle II experiment is primarily focused on the study of CP violation and baryon asymmetry, its measurements can also indirectly contribute to the understanding of dark matter. Although Belle II can not directly detect dark matter, the precise measurements of rare decays and the search for new particles and forces could provide important clues about the nature of dark matter. As a successor to the Belle experiment, which ran from 1999 to 2010, Belle II aims to collect data samples with an integrated luminosity of 50 ab^{-1} at $\Upsilon(4S)$ resonance, approximately 50 times greater than that of its predecessor. The Belle II detector, an improved version of the Belle detector, features numerous subdetector upgrades, including the addition of an entirely new subdetector for precise vertexing [1, 54].

3.1 SuperKEKB

Located at the KEK laboratory in Tsukuba, Japan, the SuperKEKB is an upgraded asymmetric-energy e^+e^- double-ring collider, following in the footsteps of KEKB. It comprises a 7-GeV electron ring (High-Energy Ring, HER), a 4-GeV positron ring (Low-Energy Ring, LER), and an injector linear accelerator (linac). A depiction of the accelerator is presented in Figure 3.1. The SuperKEKB collides electrons and positrons at a CM energy of 10.58 GeV, just above the invariant mass of the $\Upsilon(4S)$ resonance, a bound state of a bottom quark and its antiparticle ($b\bar{b}$). The $\Upsilon(4S)$ predominantly decays into two B meson pairs (B^+B^- or $B^0\bar{B}^0$) with a branching fraction exceeding 96%, thereby producing a copious number of B meson pairs. The energy differential between HER and LER is slightly less than that of KEKB (8 GeV for HER and 3.5 GeV for LER). Lowering the HER energy improves horizontal emittance and synchrotron radiation power, while increasing the LER energy reduces beam losses due to Touschek scattering [1, 54, 55].

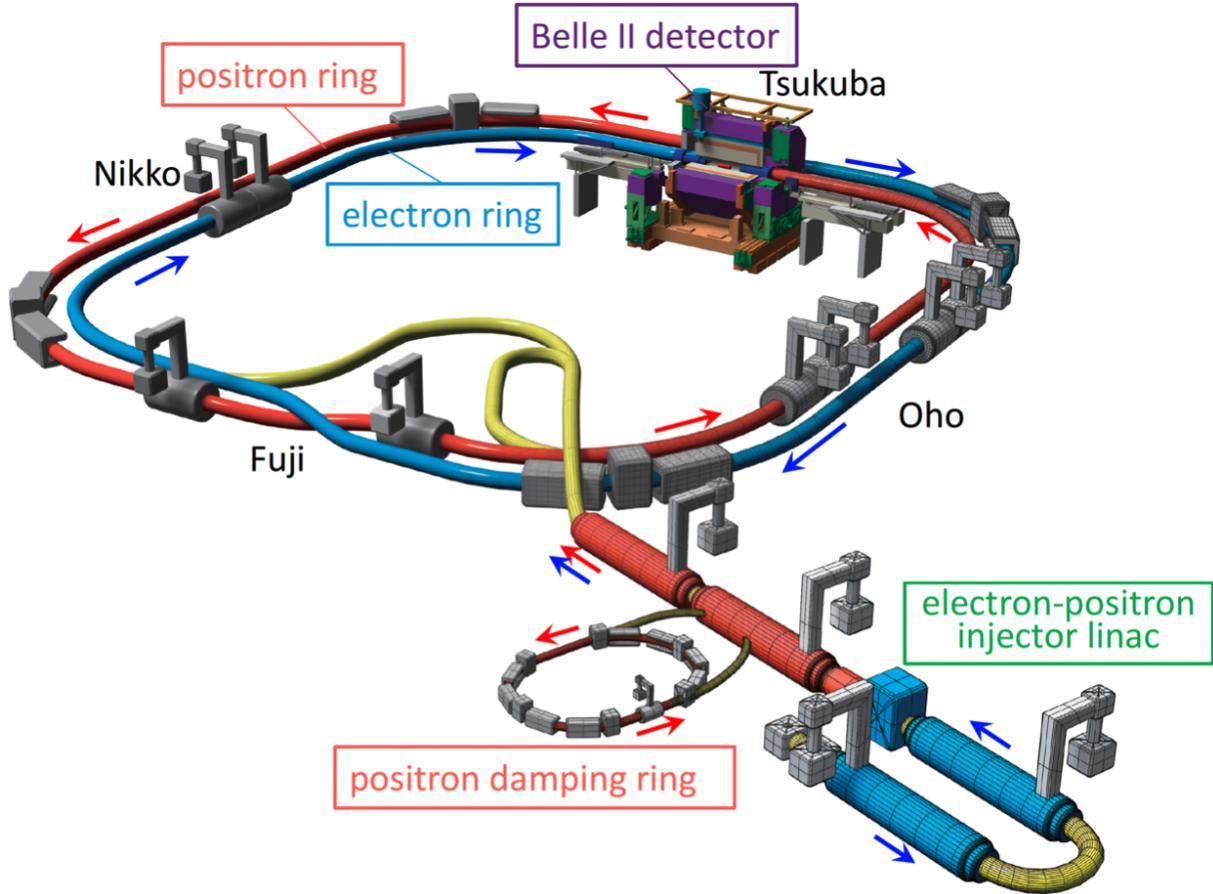


Figure 3.1: Schematic view of the SuperKEKB collider. The electron and positron beams collide at the interaction point in the Belle II detector. Image from [54].

The SuperKEKB’s asymmetric-energy design serves to perform measurements of time-dependent CP violation in the B meson system. This design boosts the β value to approximately 0.284 in the CM frame of the $\Upsilon(4S)$, resulting in a displacement of around $130 \mu\text{m}$ between the decay points of the B meson pairs, sufficient for locating the displaced vertex of the B mesons [1, 54].

The instantaneous luminosity goal for SuperKEKB is $8 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$, about 40 times greater than that of KEKB. This significant luminosity increase is expected to be achieved through a substantial reduction in beam sizes by a factor of 20 at the interaction point, according to the nano-beam collision scheme, and by doubling the beam currents in both rings [54].

Nano-beam Scheme

The nano-beam scheme is a key component of the SuperKEKB collider’s design. This scheme involves a significant reduction of the beam size at the interaction point to increase the collision rate, thereby enhancing the collider’s luminosity [1, 54].

The concept of the nano-beam scheme is rooted in the manipulation and rearrangement of the electron distribution in order to tailor the properties of the radiation. This is achieved through the use of lasers, which can create micro- and nano-structures in electron beams. The resulting beam structures can then be used to produce radiation with customizable waveforms, including fully coherent x-rays, mode-locked x-ray pulse trains, light with orbital angular momentum, and even attosecond or zeptosecond long coherent pulses in free-electron lasers [56].

The nano-beam scheme is part of a broader concept known as "beam by design" in modern accelerators. This concept involves the use of various techniques to enhance the performance of light sources through precision beam preparation, including beam conditioning, laser heating, emittance exchange, and various laser-based diagnostics [56].

3.2 The Belle II Detector

The Belle II detector with its hermetic structure is located around the Interaction Point (IP) of the SuperKEKB collider. It is composed of several sub-detectors, each designed to measure different properties of the particles produced in electron-positron collisions [1].

The first of these is the Pixel Detector (PXD). The PXD is the innermost detector of Belle II and is designed to provide precise measurements of the trajectories of charged particles close to the IP. This allows for the reconstruction of decay vertices with high precision, which is crucial for the study of B meson decays. The Silicon Vertex Detector (SVD) surrounds the PXD. The SVD also provides precise tracking information and helps in the reconstruction of decay vertices. It is particularly important for the identification of secondary vertices, which are characteristic of certain types of particle decays. The SVD together with the PXD are called Vertex Detector (VXD). The Central Drift Chamber (CDC) is the main tracking detector. It measures the trajectories of charged particles, which allows for the determination of their momenta. The CDC also provides information on the specific energy loss (dE/dx) of particles, which can be used for PID. The Aerogel Ring-Imaging Čerenkov Detector (ARICH) is one of the PID detectors of Belle II. It uses the Čerenkov effect to distinguish between different types of charged particles, such as pions and kaons. The Time-Of-Propagation Counter (TOP) is another PID detector. It measures the time of flight of particles, which, combined with their momentum information, allows for their identification. The ECL measures the energy of electromagnetic particles, such as electrons, photons, muons, and pions. It is crucial for the reconstruction of decays involving these particles. The superconducting solenoid provides a 1.5 T magnetic field that bends the trajectories of charged particles. This bending allows for the determination of the particles' momenta. Finally, the K_L^0 and Muon Detector (KLM) is designed to identify muons and neutral kaons. It is the outermost detector of Belle II and is segmented into layers of iron and scintillator. Some sub-detectors e.g. the PXD and SVD are completely new and the others e.g. the ECL are upgraded. Since the beam background is high, it is really crucial to achieve a better performance. Due to higher event rates, trigger scheme and Data Acquisition System (DAQ) are also modified [1, 55]. Figure 3.2 shows a top view of the Belle II detector, while Figure 3.3 shows a schematic view of it.

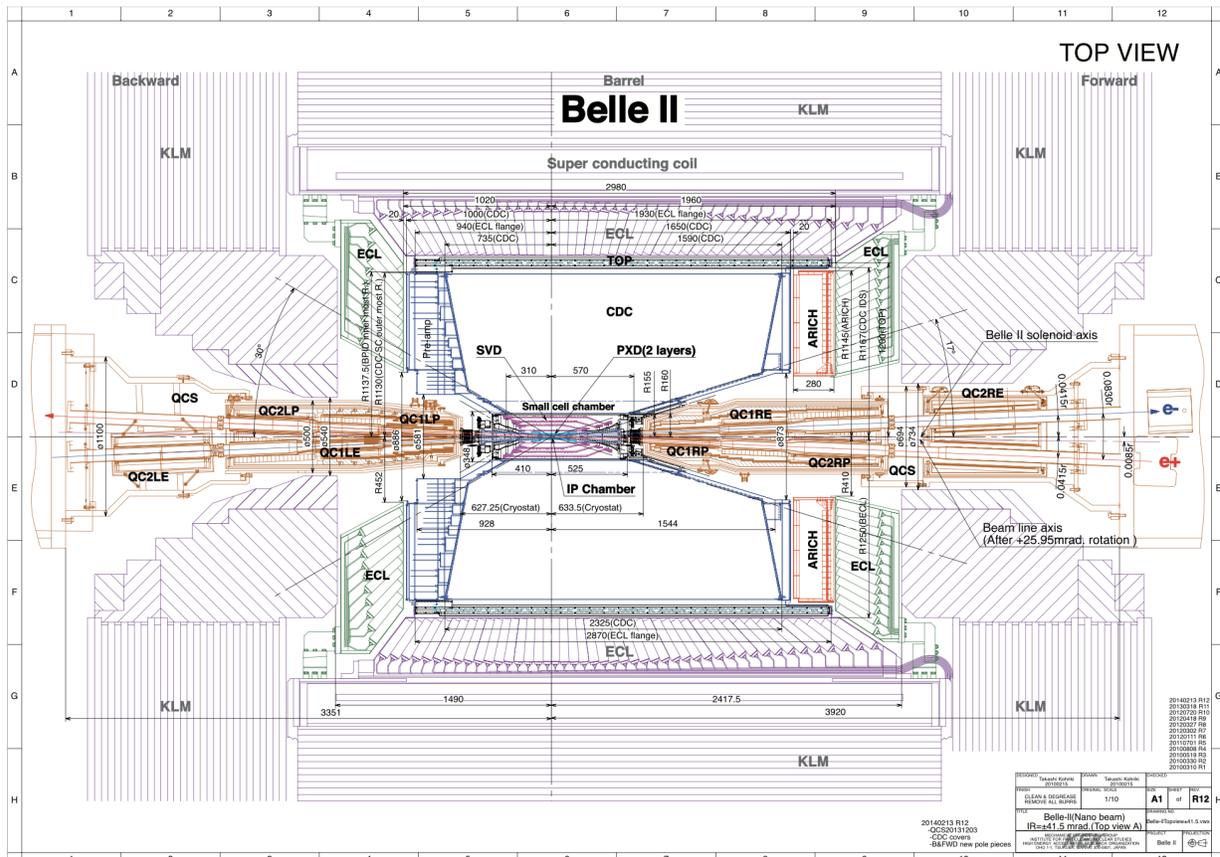


Figure 3.2: The Belle II detector top view. Image from [1].

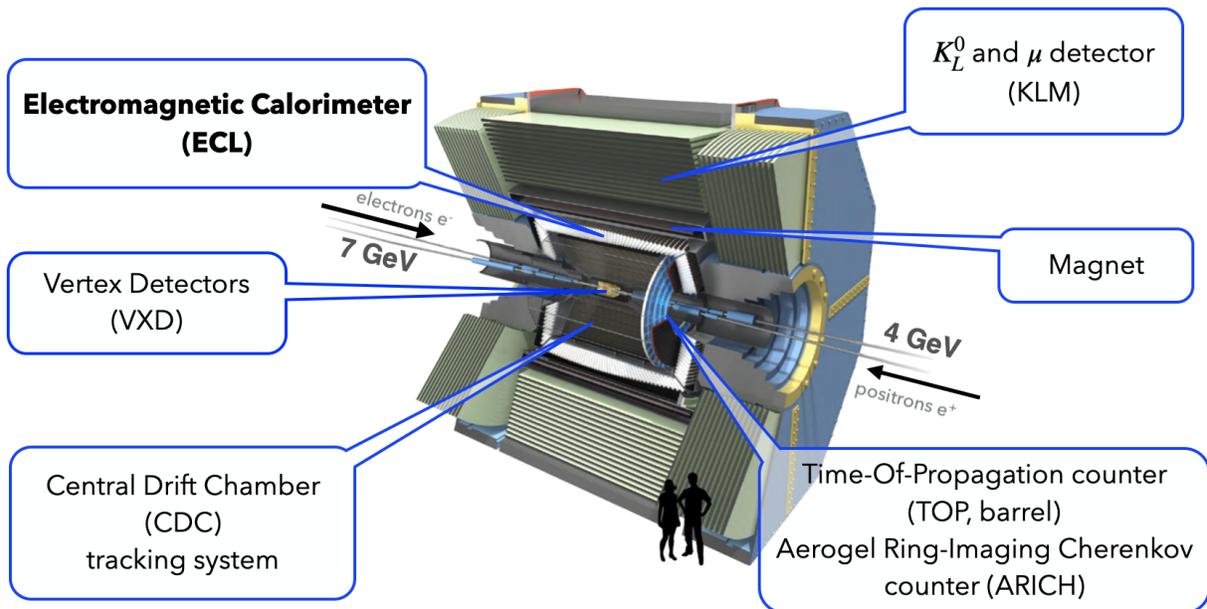


Figure 3.3: The Belle II detector schematic view. Adapted from [57].

VXD

The VXD as part of the tracking system consists of the PXD and the SVD. The VXD started at 14 mm radius and end at 135 mm including 2 layers of pixelated sensors for the PXD and 4 doubled-sided silicon strip sensors for the SVD. The hardware is completely renewed. In comparison to Belle, beam pipe, located at radius of 10 mm, and the PXD are closer to the IP, however SVD has larger radius than before. The inner PXD can measure the position of tracks with a very good spatial resolution [1, 55].

CDC

The CDC consists of 14336 sense wires arranged in 56 layers with an inner radius of 160 mm and outer radius of 1130 mm which is filled with a helium-ethane gas mixture, He-C₂H₆, with equal amounts, 50:50. It has three important tasks. First, it reconstructs charged tracks and measures their momenta accurately. Second, measurements of energy loss within the gas volume provides information concerning PID. Finally, it provides efficient trigger signals for charged particles. The information extracted from the CDC enables us to reconstruct a full 3D helix track which is essential for measuring the momentum of charged particles. The charged particles ionize the helium atoms and produce free ions and electrons. The electrons get accelerated and cascade until a measurable signal is read out at the sense wires. In order to localize the cascading ionization around the sense wires, ethane (C₂H₆) is used in the gas mixture. In comparison to Belle, number of layers and sense wires together with radius of cylinder and sense wires are all increased. However, the gas mixture and diameter of sense wires are unchanged [1, 55].

The VXD and the CDC are responsible for measuring the position and momentum of charged particles by reconstructing them in a large phase space region through the applied magnetic field of 1.5 T. The direction of magnetic field is in the direction of electron beam (+z). Both the VXD and the CDC cover the acceptance of 17° to 150° in polar angle (θ) [1, 55].

TOP

The TOP counter is located in the barrel region after the CDC and is part of the PID system. It consists of 16 quartz bar segments approximately located at radius 1200 mm. Each module has a length of 275 cm and a thickness of 2 cm. It is a special kind of Čerenkov detector where the 2D information of a Čerenkov ring image is given by the time of arrival and impact position of Čerenkov photons (γ) at photodetector at one of the end of a 2.6 cm quartz bar. The TOP counter requires photo-sensors with a single-photon time resolution of about 100 ps, which can be achieved with a 16-channel Micro-Channel Plate (MCP) Photomultiplier Tube (PMT). For identification of charged particles, the starting time i.e. particle production should be known with precision of about 50 ps [1, 55].

ARICH

The ARICH is a proximity focusing Čerenkov ring imaging detector with aerogel as Čerenkov radiator is located in the forward endcap region and is responsible to identify charged particles. It can identify and separate charged pions (π^\pm) and kaons (K^\pm) within a momentum range of approximately 0.4 GeV/c and 4 GeV/c. In order to increase the yield without degrading the Čerenkov angle resolution a new method is employed using two layers of aerogel with different refractive indices [1, 55].

The TOP and the ARICH are responsible for the PID especially for distinguishing between π^\pm and K^\pm in the barrel and forward endcap region, respectively. The TOP counter covers the acceptance of 31° to 128° in θ . Whereas the ARICH covers only the acceptance of 14° to 30° in θ [1, 55].

KLM

The KLM detector, situated beyond the superconducting coil encircling the ECL, plays a pivotal role in generating the Belle II magnetic field and is a critical component of the Belle II experiment setup. Its primary function is to discern and identify K_L^0 mesons and muons, while also serving as a magnetic flux return for the encompassing Belle II solenoid, thus enhancing the efficacy of the magnetic field within the setup. Structurally, the KLM is divided into barrel and endcap sections, each with a distinctive design tailored to maximize detection efficiency. The barrel section of the KLM is made from an arrangement of 4.7 cm thick iron plates and Resistive-Plate Chamber (RPC) superlayers. This alternating sandwich configuration not only facilitates the magnetic flux return, thereby augmenting the magnetic field's effectiveness, but also provides a substantial material density. This increased material density is instrumental in elevating the probability of interactions with K_L^0 mesons, thereby enhancing the detection and identification accuracy of these particles [1, 55].

Furthermore, the active detector elements, embedded within the layered structure, are integral to the KLM's operation. The iron plates in conjunction with the RPCs—based on glass-electrode technology—have exhibited commendable performance throughout the extensive data-collection phase of the Belle experiment. This robust performance underscores the KLM's capability in enduring large background rates, especially in certain regions of the Belle II KLM detector such as both endcaps and the innermost layers in the barrel region. These areas are anticipated to encounter elevated background rates predominantly due to neutron production, which principally arises from electromagnetic showers triggered by background reactions, like radiative Bhabha scattering. This aspect highlights the KLM's resilience and adaptability in managing background interference, thus underscoring its indispensable role in the Belle II experimental framework [1, 55].

3.3 Belle II Analysis Software Framework

The software employed in Belle II is known as basf2 [58], an acronym for the Belle II Analysis Software Framework. A vital software component, or module, within basf2 is the particle gun, which plays a key role in generating and reconstructing particle tracks. The particle gun creates particles with specified properties, including type, energy, and direction. These particles are then tracked as they move through a simulation of the detector. During the simulation, they interact with the material of the detector, generating secondary particles. The simulation results in a comprehensive prediction of the measurements that the detector would make if a particle with the specified properties were produced in an experiment. The particle gun is an essential tool for understanding the detector's response [1].

To accurately simulate the conditions of the Belle II experiment, a detector simulation is needed. The basf2 simulation package employs the Geant4 [59] software for this purpose. Short for GEometry ANd Tracking, Geant4 is a software package specifically designed to simulate the movement of particles as they pass through different materials. It is extensively used across multiple fields, including high-energy physics and nuclear experiments, and is a fundamental tool. Geant4 is written in C++ and provides all the necessary functionality for describing particle interactions with matter and managing intricate detector geometries [60].

Beam Background

With the increase in instantaneous luminosity, SuperKEKB will experience a notable rise in beam-induced backgrounds (or beam backgrounds), substantially irradiating the Belle II detector. The primary sources of beam backgrounds at SuperKEKB originate from several sources [1]:

- Touschek scattering: This phenomenon, more pronounced at SuperKEKB, involves intra-bunch scattering leading to energy deviations within a beam bunch. Measures like movable collimators and metal shields are utilized to counter this, with the background projected to be about 20 times higher than at KEKB [1].
- Beam-gas scattering: This interaction between beam particles and residual gas in the beam pipe is intensified at SuperKEKB due to elevated beam currents and a smaller beam pipe radius, expecting a Coulomb scattering rate about 100 times higher compared to KEKB. Employing vertical collimators is crucial to manage this issue [1].
- Synchrotron Radiation (SR): Originating from the beam, mainly the HER beam, SR photons exhibit energies spanning from a few to tens of keV. The inner surface of the beryllium beam pipe is coated with gold to absorb SR photons [1].
- Radiative Bhabha process: The process $e^+e^- \rightarrow e^+e^-\gamma$, which is called radiative Bhabha scattering, dominates the e^+e^- cross section at SuperKEKB. It triggers substantial neutron production impacting the outermost detectors. The rate of neutron

production correlates with luminosity, which is 40 times greater at SuperKEKB, necessitating extra neutron shielding [1].

- Two-photon process: This process corresponds to $e^+e^- \rightarrow e^+e^-e^+e^-$ creates very low momentum e^+e^- pairs that spiral around the solenoid field lines, resulting in multiple hits in the inner detectors [1].

An important aspect of the Belle II detector design is centered around maintaining high performance in spite of the increased noise and deterioration caused by beam backgrounds. These beam backgrounds expose the Belle II detector components to a substantial radiation dose from photons and neutrons, in comparison to the Belle detector, imposing a challenging operational scenario for Belle II [55].

Beam Background Overlay

This method incorporates actual background data gathered by random triggers, merging this real background event with the simulated event using digitized hits [1].

A unified framework has been developed to standardize the background overlay across all detector components. This framework comprises two basf2 modules and a base class designated for digitized hits or clusters of hits. The first module, required to operate in a single-process mode, reads data from a standard basf2 ROOT background file. The second module, capable of functioning in a multi-process mode, carries out the overlay operation [1].

The challenge of potential hit pile-up is acknowledged, with specific methods in place to address this issue. However, these methods can only approximate the pile-up since the real background data includes hits that are above the detection threshold [1].

Each class handling digitized hits is required to implement two base class methods: one to retrieve the unique channel identifier of the hit, and the other to handle the pile-up process, which is usually specific to the detector. During the overlay, the first method identifies channels where background hits are merged with the existing simulated hit. If this merging occurs, the second method is activated, determining if the pile-up criterion is met. If the criterion is not met, the background hit is added to the simulated hit collection, aiding in a more realistic simulation [1].

During this thesis, varying levels of beam background are utilized. The term $BG \times N$ ($N=0, 1, 2, 5$) in the context of the Belle II experiment refers to certain simulation conditions. The notation $BG \times 1$ denotes simulation files or conditions with nominal beam background, as opposed to $BG \times 0$ which refers to conditions without beam background. Similarly, $BG \times 2$ means there's double the nominal beam background, and $BG \times 5$ means there is five times the nominal beam background. It aids in understanding and adjusting for background noise, which can obstruct the detection of significant signal events. Beam background refers to undesired or noise events generated by the beams, unrelated to the intended beam-beam interactions [1].

3.4 Electromagnetic Calorimeter

The ECL comprises 8736 thallium-doped caesium iodide CsI(Tl) crystals situated near the interaction point, encased within a 1.5 T magnetic field, and structured with a barrel and two endcaps covering polar angles from 12.4° to 155.1° . Each 30 cm long crystal equals 16.2 radiation lengths (X_0). Illustrated in Figure 3.4, the crystals have a trapezoidal shape with front and back sides measuring approximately 5.5 cm and 6.5 cm, respectively. They are foil-wrapped, forming tower-like formations with dimensions around $5.5 \text{ cm} \times 5.5 \text{ cm} \times 30 \text{ cm}$. The barrel, containing 6624 crystals arranged in 46 rings across the θ -plane with each ring having 144 crystals in the ϕ -plane, has an inner radius of 125 cm and a 3 m length. The endcaps are positioned at roughly $z=+2 \text{ m}$ and $z=-1 \text{ m}$ with gaps for cabling and piping between them and the barrel. The gaps measure 0.8° between the Forward Endcap (FWD) and barrel, and 2° between the barrel and Backward Endcap (BWD) [1, 55, 61]. The ECL's overall layout is depicted in Figure 3.5.

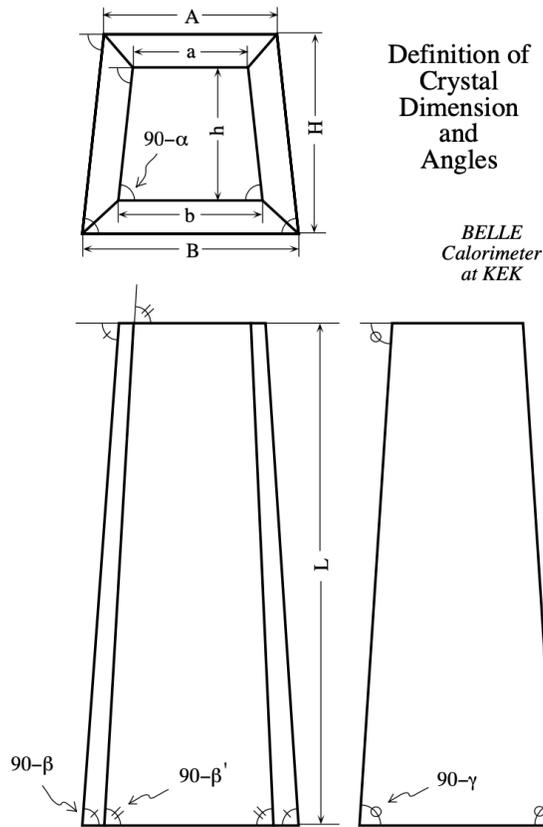


Figure 3.4: Dimension of the barrel crystals. Image from [61].

Compared to Belle, the only upgrade regarding the ECL was readout electronics and reconstruction software which was improved significantly for the latter. The new electronics considerably reduce the pile-up noise, which is essential for missing-energy studies. In the absence of background, the energy resolution is $\sigma_E/E=4\%$ at 0.1 GeV, 1.6% at 8 GeV, and angular resolution is 13 mrad at low and 3 mrad at high energies [1, 55].

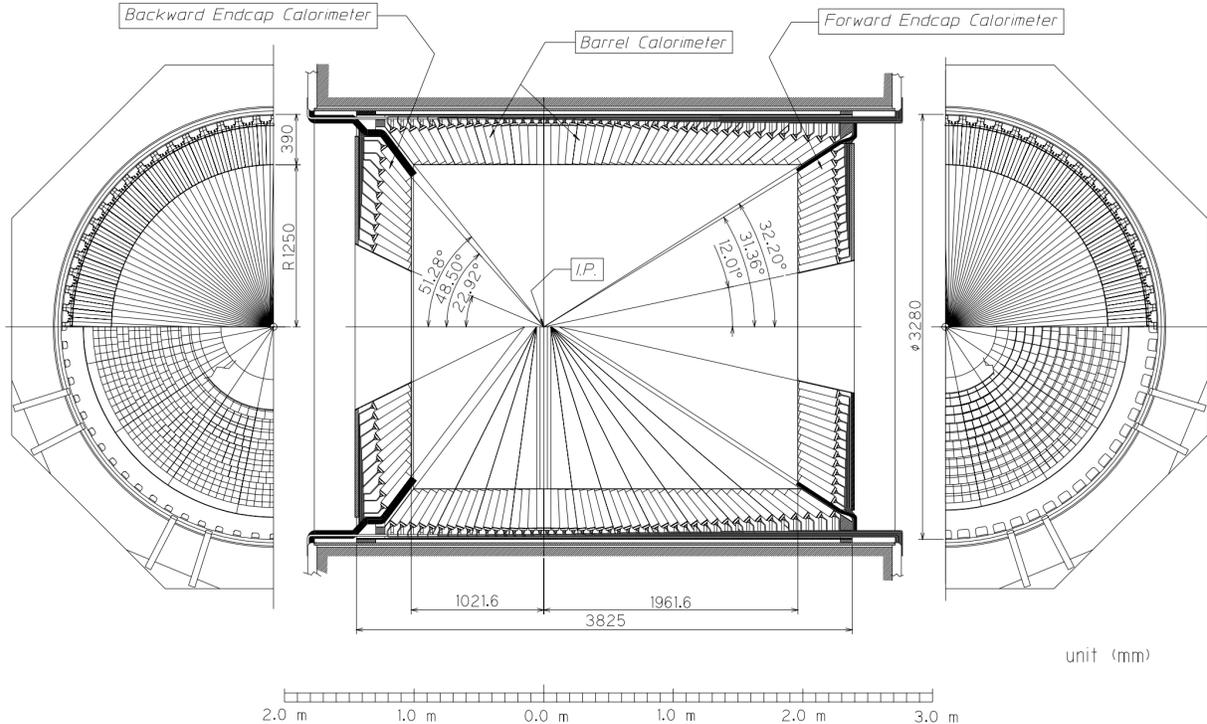


Figure 3.5: Overall design of the ECL. Image from [61].

Tracks and Clusters

In particle physics experiments, the trajectory of a charged particle is often predicted by extrapolating its track. This extrapolation is based on the information collected from signals or energy deposits, termed as **hits**, that the particle leaves in the detector's sensitive elements as it traverses through. The process of reconstructing this trajectory involves fitting the observed **hits** to a mathematical model. Depending on the context, this could be a straight line or a curve, reflecting the particle's motion in a magnetic field. An extrapolated track extends this prediction beyond the measured data points. It's crucial, especially when the detector's coverage is limited or when a particle's path deviates due to magnetic or scattering effects [62].

Within the Belle II experiment, **hits** denote the localized energy deposits captured by a detector element when a particle passes through. These **hits** are essential for reconstructing trajectories, classifying particles, and delving into the intricate physics of high-energy collisions [62].

When particles electromagnetically interact with CsI(Tl) crystals in the ECL, they produce electromagnetic showers. Each crystal has two dedicated photodiodes to detect these showers. The signals from the photodiodes are amplified and then combined to produce a waveform, which comprises 31 samples, each highlighting amplitude variations over time. A Field Programmable Gate Array (FPGA) is then used to perform a template fit on this waveform. Once calibrated, the FPGA can determine the energy transferred to the crystal by the particle and the exact time of this transfer. This data is stored in an

entity called `ECLCalDigit`, with each crystal having its unique `ECLCalDigit` [62].

Often, a single particle disperses its energy across multiple crystals. To interpret this spread, a clustering algorithm is employed. It groups crystals that are part of the same electromagnetic shower. The process starts by filtering out crystals with `ECLCalDigit` energy below a set threshold (≈ 0.1 MeV). Conversely, any `ECLCalDigit` exceeding 10 MeV is treated as a potential cluster seed. Starting from these seeds, the algorithm searches for neighboring `ECLCalDigits` and adds them to the seed if they surpass 0.5 MeV in energy. If a crystal is eligible to be added by multiple seeds, clusters merge. The search iterates, adding neighbors of newly included crystals, until no more suitable neighbors can be annexed. At the end of this procedure, each distinct cluster is labeled as a connected region [62].

ECL Variables

Extrapolated `ECLCalDigit`

From the earlier discussion on the ECL's role in PID, it is understood that in the ECL Belle II framework, the concept of the extrapolated `ECLCalDigit` is deemed significant when associating tracks with energy clusters. When a charged particle traverses the detector and leaves behind a track, the extrapolation of this track can help in predicting which calorimeter cell (crystal) the particle would interact with. This prediction is essentially what is termed as the "extrapolated `ECLCalDigit`" or `eclcaldigitExt`. It serves as a bridge between the reconstructed tracks and the energy clusters in the ECL [62].

The process begins by using the reconstructed tracks and extrapolating them to the surface of the ECL. The point where this extrapolated track intersects with the ECL provides an estimate of which calorimeter cell will likely record a signal or energy deposit from the particle. This extrapolated information is then compared with the actual recorded `ECLCalDigits` to confirm the association between a given track and its corresponding energy deposit in the ECL [62].

By effectively utilizing the extrapolated `eclcaldigitExt`, one can achieve better accuracy in associating tracks with their corresponding energy clusters. This, in turn, enhances the precision of PID and reconstruction, which is crucial for the success of the ECL [62].

Crystal positions

With the significance of the extrapolated `ECLCalDigit` established, a deeper exploration into the workings of the ECL is undertaken. Specifically, θ_{ID} is identified as a variable in `basf2` named `eclcaldigitExtCellThetaId`. It ranges from 0 in the FWD to 68 in the BWD. The numbers 14 to 57 represent the ID of a crystal in the θ -plane. The crystal numbering is shown in Figure 3.6. Crystals 0-12 belong to the FWD, 13-58 belong to the barrel, and 59-68 belong to the BWD. Each number represents a ring of crystals, with each ring containing 144 crystals in the barrel, the last two rings in the FWD (11, 12), and the first two rings in the BWD (59, 60). The number of crystals in other rings outside this range (0-10 and 61-68) is less than 144 [61].

Similarly, in basf2, there is a variable called `eclcaldigitExtCellPhiId`, denoted as ϕ_{ID} , which provides a numbering system for crystals within each ring of the θ_{ID} . The detector’s hermetic design ensures complete coverage of 360° . For the barrel section and the last two θ_{ID} rings of the FWD, as well as the first two θ_{ID} rings of the BWD, the ϕ_{ID} variable ranges from 0 to 143, resulting in a total of 144 crystals per ring [61].

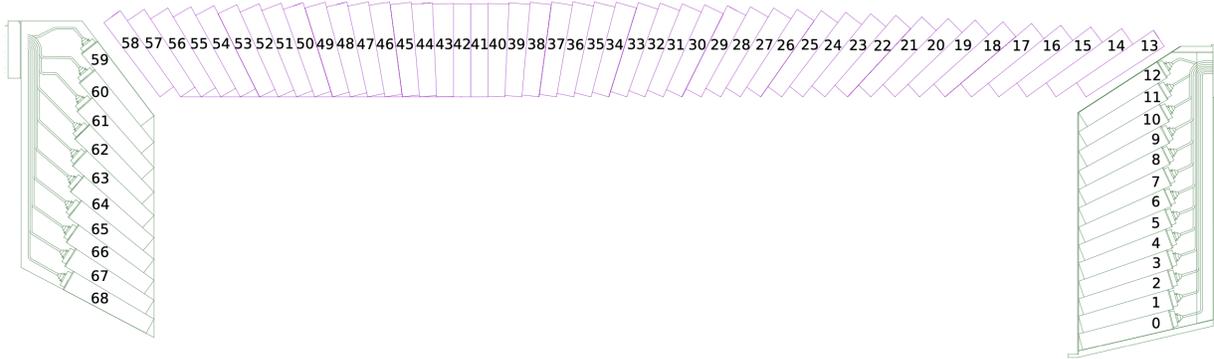


Figure 3.6: θ_{ID} of ECL crystals. Adapted from [1].

3.5 Charged Particle Identification at Belle II

Standard PID

Following a comprehensive overview of the ECL’s role in the Belle II experiment, the broader context in which it operates is considered. Specifically, a PID system for identifying charged particles is employed by the Belle II experiment, proving pivotal for various physics analyses. This PID system capitalizes on the combined strength of six sub-detectors, namely: SVD, CDC, TOP, ARICH, ECL, and KLM. These sub-detectors work in harmony, each contributing unique measurements based on a mass hypothesis i , representing particles such as electrons, muons, charged pions and kaons, protons, deuterons, and their antiparticles [1].

To elucidate further, the SVD and CDC collaborate to gauge the energy loss of particles through ionisation, represented as dE/dx . Meanwhile, the TOP and ARICH specialize in capturing optical signals that are generated relative to the velocity of charged particles. The ECL delves into analyzing light emissions from CsI(Tl) scintillation crystals, which illuminate upon energy deposition by interacting charged particles. Concurrently, the KLM focuses on determining both the depth of particle penetration and their distinct scattering patterns [63].

Building upon these foundational measures, central to this identification method is the concept of likelihood. In the context of the Belle II PID system, the likelihood, $\mathcal{L}(\mathbf{x}|i)$, gauges how well a set of parameters (associated with particle types) align with the observed data from the detectors. It essentially measures the probability of observing specific data given the parameters representing different particle types [63].

For each sub-detector, a likelihood \mathcal{L}^d is computed based on its measurements \mathbf{x}^d . These individual likelihoods are integrated to form a 'global likelihood', expressed as [63]:

$$\mathcal{L}(\mathbf{x}|i) = \prod_d \mathcal{L}^d(\mathbf{x}^d|i) = \exp\left(\sum_d \log \mathcal{L}^d(\mathbf{x}^d|i)\right). \quad (3.1)$$

Following this, the global likelihood undergoes normalization to yield the PID discriminators [63]:

$$P_i = \frac{\mathcal{L}(\mathbf{x}|i)}{\sum_j \mathcal{L}(\mathbf{x}|j)}. \quad (3.2)$$

These discriminators grant a probabilistic perspective on a charged particle's type, drawing from the entirety of the observations. To gauge the effectiveness of this identification system, specific threshold cuts on P_i are applied, spanning values between 0 and 1.

One of the distinguishing attributes of the Belle II's PID system lies in its adaptability. Users can customize the PID discriminator, choosing to rely on data from specific detectors or focusing on particular particle species. Furthermore, the system is robust enough to handle scenarios where certain detectors might not provide identification data due to reasons like geometric constraints. In such cases, a dummy value is assigned to \mathcal{L}^d for all mass hypotheses, ensuring that the overall system remains unaffected [63].

Shifting our attention more narrowly, the ECL's role in this system is of significant interest. Within the scope of this thesis, emphasis is placed solely on \mathcal{L}^{ECL} . To compare the separation power between muons and charged pions with other methods, a binary likelihood ratio is introduced as:

$$P_{\mu(\pi)}^{\text{ECL}} = \frac{\mathcal{L}_{\mu(\pi)}^{\text{ECL}}}{\mathcal{L}_{\mu}^{\text{ECL}} + \mathcal{L}_{\pi}^{\text{ECL}}}, \quad (3.3)$$

which serves as a benchmark, termed the *default*, for comparisons throughout the study.

The Belle II ECL utilizes a specialized PID algorithm. This algorithm is based on the E/p parameter, representing the ratio of the energy ascertained in the calorimeter to the momentum documented by the tracking systems. The E/p ratio plays a pivotal role in distinguishing between different particle species, especially when discerning between charged pions and muons [63].

However, challenges arise in scenarios involving particles with low momentum. For instance, the detector's solenoidal magnetic field can significantly alter the trajectory of charged pions and muons. This alteration often results in these particles navigating longer paths through various materials before they interact with the ECL. Consequently, this interaction can lead to variations in the energy detected, affecting the E/p ratio's effectiveness as a discriminator [63].

Specifically, the frequency of hadronic inelastic interactions increases at lower momenta. This surge intensifies the spread of the E/p distribution, which, in turn, considerably impedes the differentiation between charged pions and muons. A clear depiction of this phenomenon can be observed in Figure 3.7. This figure plots the E/p distribution for simulated charged pion and muon candidates situated in the central ECL barrel in two momentum regions: low and high momentum. At lower momenta ($0.2 \leq p \leq 0.6$ GeV/c), the E/p distributions for both charged pions and muons exhibit a broader spread compared to the distributions in the higher momentum region ($p > 0.9$ GeV/c). This wider spread complicates the task of distinguishing between them.

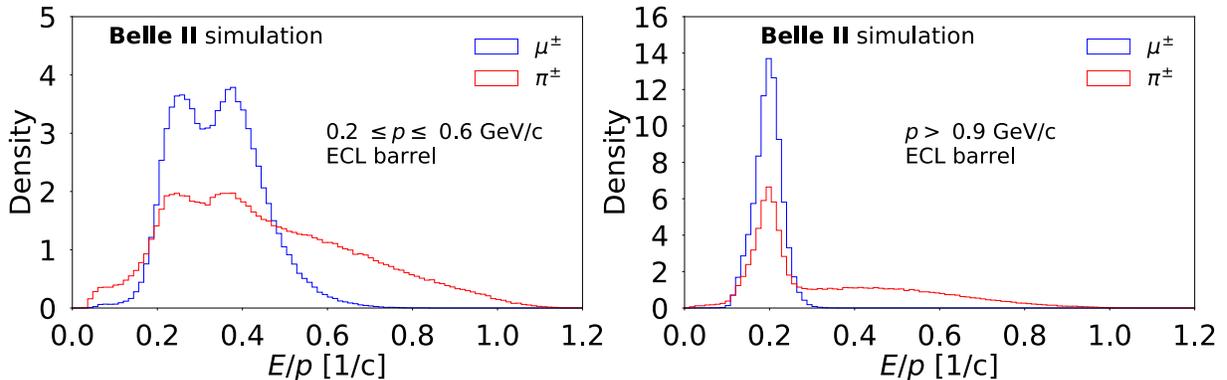


Figure 3.7: Distributions of the ratio of cluster energy over track momentum (E/p) for simulated single track candidates in the ECL barrel. The left panel shows candidates with momenta in the range of 0.2 to 0.6 GeV/c, while the right panel focuses on candidates with momenta greater than 0.9 GeV/c. Muon candidates are represented in blue, and charged pion candidates are shown in red.

To explain the double-peak behaviour observed in muon energy distributions, one must delve into the details of the ECL clustering algorithm. This algorithm has been designed with a tendency towards forming clusters that are radially symmetric. The central point of these clusters is the crystal that captures the maximum energy, commonly known as the seed crystal. Such a configuration is especially adept at capturing the energy deposition patterns characteristic of photons and electrons. These particles usually deposit their energy in a concentrated, symmetric fashion, aligning well with the algorithm’s design [63].

However, when this algorithm encounters minimally-ionising particles, notably muons, the symmetry breaks. Muons, due to their unique interaction dynamics, tend to ionize the medium more sparsely. This results in a more dispersed energy deposition. Consequently, rather than being confined to a singular symmetric cluster, the energy of muons often gets fragmented across multiple clusters. This fragmented energy deposition gives rise to the double-peak or bimodal distribution observed in Figure 3.7 (left plot). Such differences highlight the complexities in PID and emphasize the significance of understanding the detector design and its data interpretation implications [63].

Boosted decision trees (BDT)

While standard PID method offer insight into PID, the complexities and intricacies involved in the ECL interactions require more sophisticated techniques. This leads to the exploration of multivariate approaches like BDTs. In the Belle II software, the development of particle showers is characterized as a key challenge. Various metrics, like the Zernike moments [64], help capture the lateral progression of these showers. Additionally, the depth to which extrapolated tracks penetrate the ECL, denoted as ΔL provides valuable insights into the interaction dynamics of different particles [65]. For instance, while electrons typically interact via electromagnetic interaction, muons are minimally ionising. On the other hand, hadrons, though also minimally ionizing, can experience strong inelastic interactions with the ECL material, causing what are known as hadronic split-offs. The longitudinal development of these particle showers, or how they progress over distance, is another crucial aspect captured by metrics such as ΔL [63].

Given the intricate nature of these interactions and the vast amount of information available, a multivariate approach is adopted for PID within the ECL. Traditional methods might struggle with the high dimensionality and inter-correlations among these observables. Enter BDTs, a ML technique adept at handling such complexities. BDTs can decipher and exploit subtle relationships among inputs, leading to superior particle classification performance. Specifically, the Belle II ECL software employs gradient boosting decision trees from the TMVA [66] package. This approach is tailored for binary classifications, distinguishing between electrons and hadrons, as well as between muons and hadrons [63].

These shower shape variables are [63]:

- E/p : Ratio of cluster energy over track momentum.
- E_{cluster} : Cluster energy.
- E_1/E_9 : Ratio of the energy of the seed crystal over the energy sum of the 9 surrounding crystals.
- E_9/E_{21} : Ratio of the energy sum of 9 crystals surrounding the seed over the energy sum of the 25 surrounding crystals (minus 4 corners).
- $|Z_{40}|$: Zernike moment $n = 4$, $m = 0$, calculated in a plane orthogonal to the EM shower direction.
- $|Z_{51}|$: Zernike moment $n = 5$, $m = 1$, calculated in a plane orthogonal to the EM shower direction.
- Z_{MVA} : Score of BDT trained on 11 Zernike moments.
- ΔL : Projection on the extrapolated track direction of the distance between the track entry point in the ECL and the cluster centroid.
- $\Delta \log \mathcal{L}(\ell/\pi)_{\text{CDC}}$: Log-likelihood differences between lepton and pion in the CDC.

- $\Delta \log \mathcal{L}(\ell/\pi)_{\text{TOP}}$: Log-likelihood differences between lepton and pion in the TOP.
- $\Delta \log \mathcal{L}(\ell/\pi)_{\text{ARICH}}$: Log-likelihood differences between lepton and pion in the ARICH.
- $\Delta \log \mathcal{L}(\mu/\pi)_{\text{KLM}}$: Log-likelihood differences between muon and pion in the KLM.

Since the above-mentioned observables are highly correlated, BDTs can understand non-trivial dependencies across inputs to improve classification performance [63].

Chapter 4

Convolutional Neural Networks

ML, a subfield of artificial intelligence, has its roots in the mid-20th century, with the development of the perceptron in the late 1950s. Over the years, the field has evolved to include various types of algorithms, ranging from decision trees to support vector machines and neural networks [67]. The primary aim of ML is to enable computers to learn from data, thereby automating analytical model building.

Supervised learning is an ML paradigm where a model is trained on a dataset containing labeled examples, each paired with the desired output value. During training, the model learns to associate the input data with the correct output, making predictions and gradually adjusting its weights based on the difference between its predictions and the actual values, a process typically facilitated through a loss function. The aim is to fine-tune the model to a point where it can make accurate predictions when presented with new, unseen data. Common applications include regression tasks such as predicting house prices, and classification tasks like identifying spam emails or recognizing handwritten digits. The performance of a supervised learning model is evaluated by measuring its accuracy on a separate set of labeled data called the test dataset. The requirement of a large, well-labeled dataset is both a strength and a limitation of supervised learning, as quality data can lead to highly accurate models, but collecting and labeling this data can be expensive and time-consuming [67, 68].

CNNs were inspired by the visual cortex and were initially developed for image recognition tasks. Yann LeCun's work in the late 1990s, particularly on the LeNet architecture, was seminal in demonstrating the capabilities of CNNs [69]. Since then, CNNs have become the go-to architecture for any task related to image perception, achieving state-of-the-art results in various domains.

CNNs are particularly adept at automatically and adaptively learning spatial hierarchies of features. This makes them well-suited for analyzing pixel images of energy depositions in the Belle II ECL [2]. Their ability to learn from the spatial configuration of data makes them more effective than traditional ML algorithms for this specific application [69].

4.1 Convolution

Convolutional networks have achieved significant success in real-world uses. The term CNN suggests that the network utilizes a specific mathematical process known as convolution. This process is a unique type of linear function. Essentially, CNNs are neural networks that replace general matrix multiplication with convolution in one or more of their layers. The convolution operation is the cornerstone of CNNs. Mathematically, it is defined as [68]:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(i - m, j - n)K(m, n). \quad (4.1)$$

Here S is the output feature map, I is the input image, and K is the kernel or filter. The kernel slides over the input image, performing element-wise multiplication and summing the results to produce each pixel of the output feature map [68]. The process is visualized in Figure 4.1. This visualization illustrates how a 3×3 filter is applied to the input data, a fundamental operation in the convolution process. The filter traverses the entire spatial extent of the input data, performing element-wise multiplications between the filter and the overlapping section of the input data. The results of these multiplications are then summed to form a single output pixel in the resulting feature map. This operation is systematically performed across the entirety of the input space, yielding a feature map that reveals the locations of features, as identified by the filter, within the input.

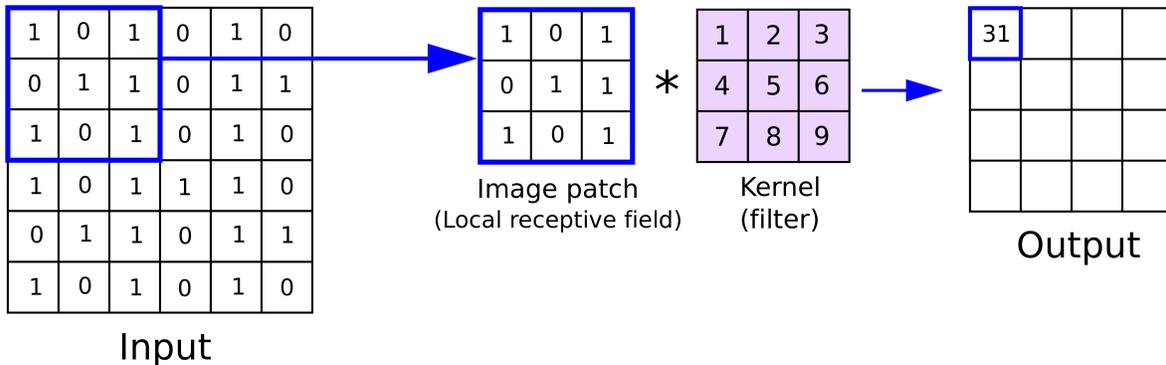


Figure 4.1: Illustration of the convolution operation in CNNs. Image from [70].

Traditional neural network layers involve matrix multiplication where each input unit interacts with every output unit through individual parameters. In contrast, convolutional networks often have limited interactions, known as sparse connectivity or weights. This is achieved by having a kernel size smaller than the input. For instance, in image processing, while the input image may have thousands or even millions of pixels, small yet significant features like edges can be detected using kernels that cover only a few tens or hundreds of pixels. This results in reduced memory needs for storing parameters, enhancing the model's statistical efficiency and requiring fewer operations for output computation. The efficiency gains from this approach are typically substantial [68].

The convolution operation allows the network to focus on local regions of the input, learning to recognize basic shapes or patterns like edges and corners in the first layer, and more complex structures in deeper layers. This hierarchical feature learning is what makes CNNs highly effective for image-based tasks [68].

Feature Maps

The feature map is the output of the convolution operation. It represents the spatial arrangement of features detected by the kernel in the input image. For example, a kernel designed to detect vertical edges would produce a feature map highlighting the vertical edges in the image. Feature maps serve as the new "images" that are passed to subsequent layers. As we go deeper into the network, these feature maps represent increasingly abstract features. By the time we reach the fully connected layers, these feature maps are rich representations that capture the high-level content of the input, making it easier for the network to perform tasks like classification [68].

4.2 Fully Connected Layers

Fully connected layers are typically used towards the end of the network. In these layers, each neuron is connected to every neuron in the previous layer. The primary purpose of the fully connected layers is to perform classification based on the high-level features extracted by the preceding convolutional layers [67]. In a fully connected layer, the output y is related to the input x as:

$$y = Wx + b. \tag{4.2}$$

Here, W is the weight matrix and b is the bias vector [68].

As shown in Figure 4.2, a typical CNN is composed of multiple layers, including convolutional, pooling, and fully connected layers. This representation visually outlines the progression of data through the network, from raw input data to the final class scores after passing through the various layers.

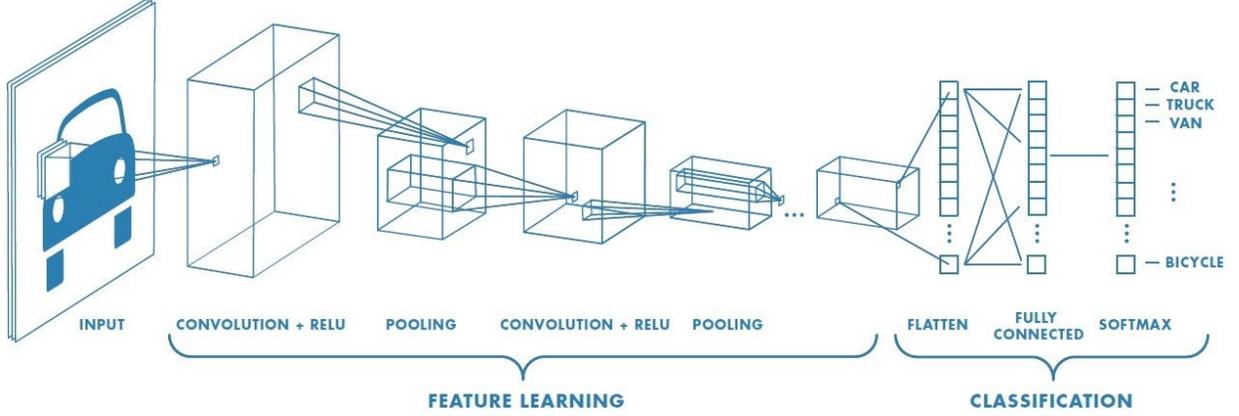


Figure 4.2: A typical CNN architecture showing convolutional, pooling and fully connected layers. Image from [71].

4.3 Embedding

Embedding is a powerful technique in deep learning that transforms discrete categorical variables, such as the θ_{ID} and ϕ_{ID} of the crystals in the ECL Belle II , into continuous vectors. This transformation facilitates the processing of such data by neural networks, which require numerical input. By mapping high-dimensional data into a lower-dimensional space, embeddings preserve the semantic relationships between data points [68].

In the context of CNNs, embeddings capture spatial hierarchies and relationships, making them particularly effective for tasks that rely on the spatial configuration of the crystals in the ECL Belle II .

Mathematically, an embedding layer functions as a lookup table. Given an input D , such as θ_{ID} or ϕ_{ID} , the embedding layer returns the corresponding vector:

$$E : D \rightarrow \mathbb{R}^d. \tag{4.3}$$

Where E is the embedding function and d is the dimensionality of the embedding space. These embedding vectors, initialized with small random values, are refined during training through backpropagation to minimize the loss [68].

When fed into a CNN, embeddings of θ_{ID} and ϕ_{ID} enable the network to discern spatial hierarchies and features, enhancing its performance in tasks like energy deposition prediction or anomaly detection in the crystals.

In summary, the combination of embeddings and CNNs provides a robust approach for processing structured data in applications like the ECL Belle II , where spatial configurations are crucial.

4.4 Activation Function

Activation functions introduce non-linearity into the network. Without non-linearity, no matter how many layers the network has, it would behave just like a single-layer model because the composition of linear functions is itself a linear function [68].

The most commonly used activation function is the Rectified Linear Unit (ReLU), defined as:

$$\text{ReLU}(x) = \max(0, x). \quad (4.4)$$

ReLU is computationally efficient and helps mitigate the vanishing gradient problem, which is crucial for deep networks [68].

4.5 Preprocessing

Preprocessing is a pivotal stage in the ML pipeline, tailored to transform data into an ideal format and structure for model training. Various preprocessing techniques exist, and their application varies depending on the specific problem at hand [67].

- **Normalization:** One of the most common preprocessing techniques is normalization. By scaling input features to have zero mean and unit variance, normalization ensures that all features contribute equally to the model's performance. This is especially crucial for models that rely on gradient-based optimization algorithms, as features with larger scales can disproportionately influence the model's learning [67].
- **Data Augmentation:** For image data, data augmentation is a powerful technique to artificially increase the size of the training dataset. By applying transformations such as rotations, translations, zooming, and flipping, models can be trained on a more diverse set of examples, leading to better generalization. Augmentation not only helps in improving the model's robustness but also in preventing overfitting [67].
- **Resizing:** Resizing images to a standard size ensures that they can be processed by the model in a consistent manner. This is particularly important for CNNs, which often expect input images of a fixed size [67].
- **Other Techniques:** Other preprocessing steps might include handling missing data, encoding categorical variables, and feature extraction. The choice of preprocessing techniques often depends on the nature of the data and the specific problem being addressed [67].

4.6 Datasets: Training, Validation, and Test

In ML, it's essential to evaluate the model's performance in various scenarios to ensure its robustness and reliability. Therefore, a dataset is divided into three independent datasets [72]:

- Training dataset: The training dataset is used to train the model. It's the primary dataset on which the weights of the model are adjusted during the learning process.
- Validation dataset: The validation set plays a pivotal role in model selection and hyperparameter tuning. After each epoch (or a set number of iterations) during training, the model's performance is evaluated on the validation set. This helps in monitoring the model for overfitting and deciding when to stop training, a technique known as early stopping.
- Test dataset: After the model has been trained and the best hyperparameters have been selected, it's evaluated on the test set. This dataset provides an unbiased evaluation of the model's generalization performance on new, unseen data.

Overfitting and Underfitting

Overfitting occurs when the model becomes too complex and starts to memorize the training data rather than generalizing from it. While the performance on the training data might be excellent, the model performs poorly on unseen data. Techniques like regularization, dropout, and increasing the amount of training data can help mitigate overfitting [72].

Underfitting is the opposite of overfitting. It occurs when the model is too simple to capture the underlying patterns in the data. This results in poor performance on both the training and test data. Increasing the model's complexity, adding more features, or using a different model architecture can help address underfitting [72].

4.7 Backpropagation

Backpropagation serves as the cornerstone of learning in neural networks, enabling the model to refine its predictions by minimizing the error between the predicted and actual outputs. This supervised learning algorithm computes the gradient of the loss function with respect to each weight, leveraging the chain rule of differentiation [73]. Mathematically, it's expressed as:

$$\frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \frac{\partial y}{\partial w}, \tag{4.5}$$

where L represents the loss function.

Originating from the manner in which errors are propagated backward through the network, the term backpropagation aptly describes the process. Starting from the output layer, errors are relayed backward to the input layer. At each juncture, weights are adjusted based on the computed gradient, with the learning rate determining the step size taken in the direction of the negative gradient during these updates. An appropriately selected learning rate ensures that the model converges towards the minimum of the loss function, thereby optimizing its performance [67, 68, 72, 73].

Loss Function

The loss function L quantifies how well the model's predictions \hat{y} match the true labels y . It serves as the objective function that the training process aims to minimize [67].

For classification problems, Cross-Entropy is often used:

$$L = - \sum_i [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (4.6)$$

For regression problems, Mean Squared Error (MSE) is commonly used:

$$L = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (4.7)$$

Both of these loss functions are differentiable, which is a necessary property for optimization using gradient-based methods [67].

Gradient Descent and Optimizers

Gradient descent is an optimization algorithm used to minimize the loss function L . The weights \vec{w} are updated in the opposite direction of the gradient $\vec{\nabla}L$:

$$\vec{w}' = \vec{w} - \alpha \vec{\nabla}L. \quad (4.8)$$

Here, α is the learning rate, which controls the size of the steps taken during the optimization process [68]. To grasp this idea more clearly, picture a ball descending a valley's slope as shown in Figure 4.3. Here, the valley represents the form of the loss function. When the ball is placed in the valley, it naturally moves towards the bottom. Initially, we pick a random point for the ball to start, and then visualize its journey downwards. This movement can be represented using the derivatives of the loss function. In fact, the second derivative reveals the contour of the valley, indicating if it's a saddle point or a deep pit [74].

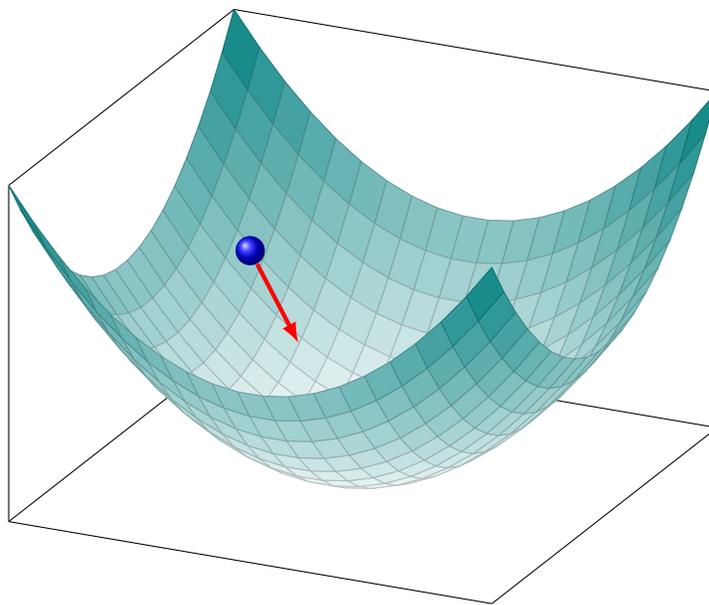


Figure 4.3: The problem of finding a global minimum by rolling a ball down the valley.

Stochastic Gradient Descent

In Stochastic Gradient Descent (SGD), the update is performed for each data point, rather than for the entire dataset. This introduces randomness, which can help escape local minima but also introduces noise into the optimization process [75].

More advanced optimizers like Adam [76] combine the advantages of both AdaGrad and RMSProp. Adam computes adaptive learning rates for each parameter and also uses momentum by adding a fraction of the previous gradient to the current one.

4.8 Hyperparameters

In the realm of ML, particularly when dealing with neural networks, hyperparameters play a crucial role in determining the behavior and performance of the model. Unlike parameters, which are learned during training, hyperparameters are set before the training process begins and guide the learning process [68].

Key hyperparameters include the batch size, number of epochs, and learning rate. The batch size dictates the number of samples utilized to update the model in each iteration. A smaller batch size can offer a regularizing effect and lower generalization error, but it might also lead to a less stable convergence. On the other hand, the number of epochs defines how many times the learning algorithm will iterate over the entire training dataset. Too few epochs can result in underfitting, while too many might lead to overfitting. Lastly, the learning rate, often denoted as α , controls the step size taken during the optimization process. A high learning rate might overshoot the optimal solution, while a low rate might converge too slowly or get stuck in local minima [68, 74].

For CNNs, additional hyperparameters related to the convolution operation come into play, such as kernel size, stride, padding, and dilation. The kernel size determines the dimensions of the kernel, essentially defining the receptive field of the convolution. Stride affects how the kernel moves across the input image, influencing the spatial dimensions of the output. Padding, by adding extra pixels around the input image, ensures spatial dimensions are preserved post-convolution. Dilation, less commonly used, adjusts the spacing between pixels in the kernel, allowing for larger receptive fields without increasing the kernel size [67, 77].

Given the high-dimensional space of hyperparameters, especially for complex models like CNNs where even the input features, such as $N \times N$ pixel images, play a role, finding the optimal set becomes a challenge. Techniques like grid search, where a range of values is pre-specified for each hyperparameter, or random search, where values are sampled from predefined distributions, can be employed to navigate this space and identify the best combination [78].

4.9 Regularization

Regularization techniques are essential tools in the ML toolkit, aiming to prevent overfitting and enhance the generalization capability of models. One such technique is dropout. During training, dropout randomly drops a subset of neurons, ensuring that no single neuron becomes overly specialized. This introduces a form of noise into the training process, making the model more robust and less prone to overfitting [68, 79].

Another regularization strategy is early stopping. Here, the training process is halted before convergence if there's no improvement (or even degradation) in the model's performance on a held-out validation dataset over a specified number of epochs. The patience parameter determines how many epochs without improvement should be tolerated before stopping. This approach ensures that the model doesn't overtrain on the training data, preserving its ability to generalize well to new, unseen data [68, 80].

Chapter 5

Particle Identification with the CNN

In the preceding chapters, an understanding was established regarding the theoretical background of physics processes as particles traverse through detector materials. The crucial role of PID and the advantages of μ/π separation were highlighted, alongside an introduction to ML techniques, with a focus on CNNs. This chapter explains the application of CNNs for the crucial task of separating muons and charged pions within the Belle II ECL.

The ensuing sections will discuss the processes of event generation, track reconstruction, and the criteria for selection. This will be followed by a comprehensive overview of data preprocessing, model training, and test procedures. Subsequent to this foundational groundwork, a comparative analysis of the new methodology against other PID techniques in the Belle II ECL will be undertaken. Additionally, explorations into the implications of varying thresholds and the influence of beam background will be presented.

5.1 Event Generation and Reconstruction

The process of training neural networks for the Belle II ECL begins by generating samples to feed the networks. The samples discussed in this chapter are MC simulation for single type of particle per dataset.

Using particle gun, approximately 2 million single muon and charged pion candidates are generated separately. For each particle type, μ^+ , μ^- , π^+ , π^- , 500 000 events are generated using basf2. The samples are generated with beam background overlays $BG \times 1$ simulated for early data taking conditions. Each track is first reconstructed in the tracking detectors and then extrapolated into the ECL. Muons and charged pions are generated with the following conditions:

- Number of tracks: In order to obtain clean samples of training dataset, one track per event is generated.
- Transverse momentum: The transverse momentum (p_T) is uniformly generated with a flat distribution between 0.2 GeV/ c and 1 GeV/ c . The lower range of p_T is chosen

slightly below the threshold at which a track can reach the ECL, $\approx 0.28 \text{ GeV}/c$. However, this does not mean that particles with lower momentum, and hence lower energy, cannot reach the ECL. If tracks bend sufficiently, they partially reach the ECL and deposit energy in the crystals. The upper limit of p_T is set higher than the minimum limit for a track to reach the KLM, $\approx 0.7 \text{ GeV}/c$.

- Polar angle: The polar angle (θ) is uniformly generated with a flat distribution between 27° and 134° . To gather more information regarding the passage of tracks through small gaps between the ECL barrel and FWD, and the ECL barrel and BWD, the θ range is selected to be slightly larger than the ECL barrel acceptance, which is between 32.2° and 128.7° .
- Azimuthal angle: The azimuthal angle (ϕ) is uniformly generated with a flat distribution between 0° and 360° .
- The track is generated from the IP with coordinates $(0, 0, 0)$ cm. Particles coming from outside of the IP are not part of the scope.

5.2 Event Selection

To select suitable samples for the training dataset, muon and charged pion candidates must meet additional conditions. After the generation and reconstruction of candidates, the following selection conditions are applied:

- The polar angle ID (θ_{ID}) of the crystal hit by an extrapolated track falls within the ECL barrel.
- Generated tracks are truth-matched using MC information.
- In each event, track with the highest p_T is selected. Although each event is expected to have only one reconstructed track, there are cases ($< 2\%$) with two reconstructed tracks.

This study focuses on the ECL barrel as it enables the generation of symmetrical pixel images. When a track passes through the detector and reaches the ECL, a window of 7×7 crystals centered around the crystal hit by the track is selected. As the track travels through the ECL, energy is deposited and stored due to interactions with the detector material, resulting in energy loss. To ensure the creation of symmetrical 7×7 pixel images, the central crystal, represented by θ_{ID} , must have a crystal index value between 14 and 57. It is important to note that pixel images with θ_{ID} outside this range are not symmetrical due to the varying number of crystals involved.

Even though the tracks are generated with a lower limit of $0.2 \text{ GeV}/c$ for p_T , the reconstructed p_T distributions consistently show a noticeable minimum value around $0.28 \text{ GeV}/c$.

This minimum p_T represents the approximate threshold for a charged track to reach the ECL barrel, as discussed earlier.

The dips observed in the reconstructed θ distributions around 32° and 128° are a result of the gaps between the ECL barrel and the FWD, as well as the BWD.

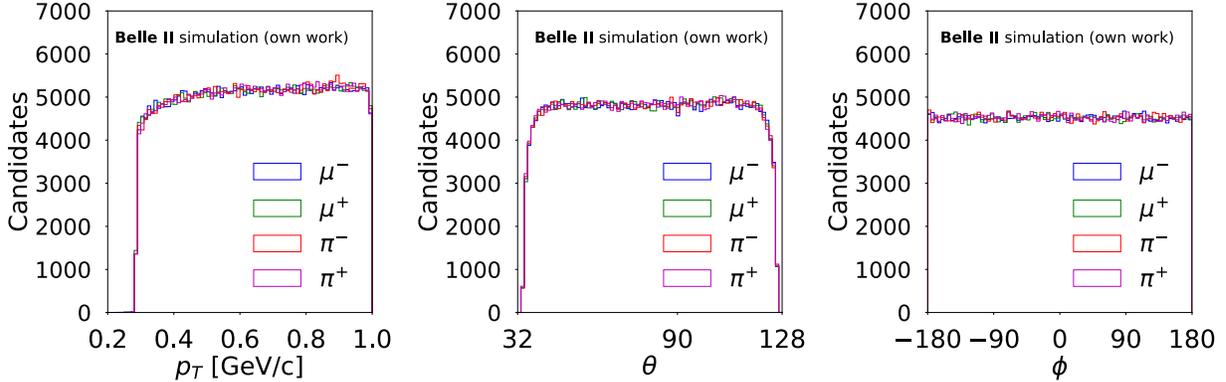


Figure 5.1: Distributions of reconstructed transverse momentum (p_T), polar angle (θ), and azimuthal angle (ϕ) (from left to right) for events generated under $BG \times 1$.

5.3 Inputs and Preprocessing

As the journey into CNNs progresses, it becomes evident that the quality and characteristics of the inputs, paired with the preprocessing steps, are pivotal cornerstones. Originally designed for image and video data, CNNs demand preprocessing to achieve their optimal performance. This encompasses several stages, from image resizing to dataset augmentation. Each step is meticulously crafted to reduce computational burdens, ensure data variations are accounted for, and fortify model resilience [81, 82, 83]. A testament to the importance of preprocessing is the widely accepted notion that data preparation, including preprocessing, consumes 80% of the total time in data science projects, leaving only 20% for model creation and evaluation. Thus, the intricate dance of preprocessing - involving data cleansing, handling missing values, numeric value normalization, and categorical variable encoding - is not just a preliminary step, but a cornerstone for ensuring precise and meaningful pattern detection by machine learning algorithms.

Input Handling

Within this study, CNN models ingest inputs at two distinct junctures: prior to and after convolution. One input stream encompasses the energy deposition in 7×7 pixel images, while the other channel encompasses extrapolated track parameters, including p_T , θ_{ID} , and ϕ_{ID} . The latter parameters are introduced post convolutional layers. As previously discussed, θ_{ID} and ϕ_{ID} are integers indicative of the crystal's location (index) within the ECL. Energy values in pixel images remain unaltered and unnormalized, given

their inherently modest magnitudes. However, some pixel images portraying charged pions show elevated energy values (> 1 GeV), arising from pion inelastic interactions with nuclei, resulting in proton production. To mitigate the impact of these instances, the extreme values are replaced by 1 GeV. However, these replacements are isolated occurrences and exercise minimal influence over the mean and standard deviation of energy depositions.

Figure 5.2 furnish visualizations of typical energy deposition patterns for muons and pions within 7×7 pixel images. The upper row pertains to muons (blue), whereas the lower row represents pions (red). Each pixel corresponds to a crystal with an area of approximately 5×5 cm², culminating in an image scope of around 35×35 cm². The x -axis signifies the θ_{ID} orientation, while the y -axis signifies ϕ_{ID} . Notably, muons show more localized energy deposition patterns compared to pions. Nonetheless, a comprehensive evaluation encompassing a larger sample size is imperative to affirm the viability of this pattern for μ/π separation.

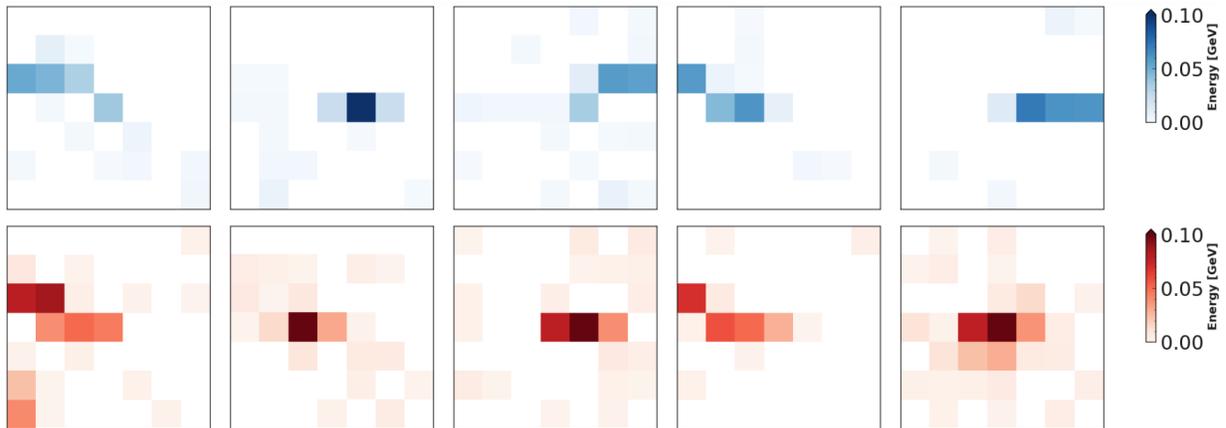


Figure 5.2: Typical 7×7 pixel images of muons (blue) and pions (red) in the ECL barrel. The color intensity represents the energy deposition, with deeper colors indicating higher energies.

Table 5.1 displays key statistics - minimum, maximum, mean, and standard deviation - across all these images, offering a glimpse into the energy distribution for different particle types and pixel configurations across various scenarios. It helps to grasp the range, central tendencies, and variability of energy deposition within these conditions. An important observation from this table is that high-energy pixel images are associated with pions, and upon closer inspection, it is found that such pixels predominantly occur on the edges of the images. This suggests that higher energy values in pixels tend to group around the edges due to pion inelastic interactions with nuclei.

The focus on 7×7 pixel images in this study is justified based on factors like maintaining performance levels, ensuring adequate coverage (particularly for symmetric pixel images), and managing noise sensitivity. Larger dimensions in $N \times N$ pixel images could possibly result in reduced performance and narrower θ_{ID} coverage, with the issue of increased noise levels becoming more significant, especially in the context of the Belle II experiment where higher beam background is expected. Moreover, the closeness of tracks being studied to

other tracks within larger $N \times N$ pixel images could further deteriorate performance due to high-energy pixels.

The 7×7 pixel images alongside $BG \times 1$ are chosen for the final stages of training and testing, as $BG \times 1$ samples align better with the real-world conditions of the Belle II experiment. However, rigorous testing with higher beam backgrounds is also conducted to evaluate performance of CNN models against these conditions which will be presented in later sections.

Table 5.1: Energy statistics in all pixels of 7×7 images for events generated under $BG \times 1$.

		μ^-	π^-	μ^+	π^+
Energy (GeV)	minimum	0.000	0.000	0.000	0.000
	maximum	0.537	3.505	0.627	5.692
	mean	0.004	0.006	0.004	0.007
	standard deviation	0.016	0.026	0.015	0.030

As mentioned earlier, the variables reconstructed p_T , θ_{ID} , and ϕ_{ID} are input parameters provided after the convolutional layers. Figure 5.3 displays the distributions of these variables for samples with $BG \times 1$. To ensure that the tracks can reach the ECL, a minimum cut of $0.28 \text{ GeV}/c$ is applied to the reconstructed p_T distribution.

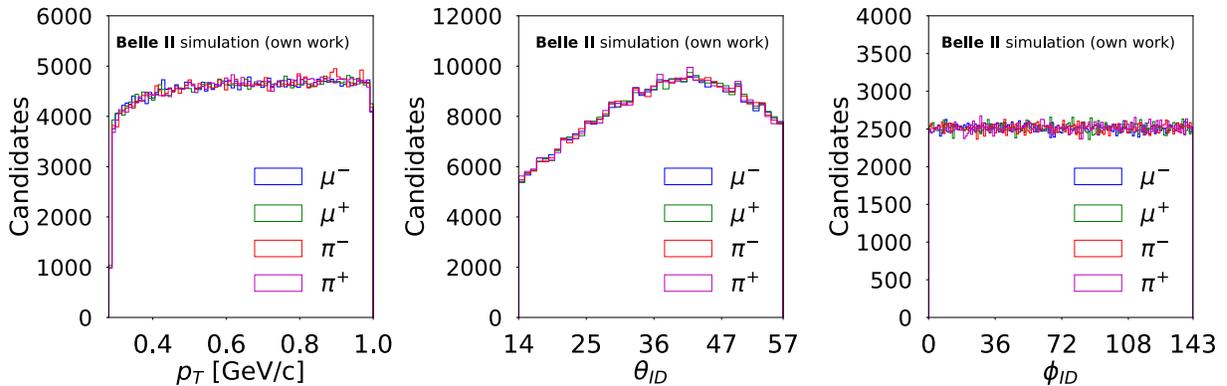


Figure 5.3: Distributions of reconstructed p_T , θ_{ID} , and ϕ_{ID} for events generated under $BG \times 1$

Data Preprocessing

The first part of the data preprocessing has already been addressed in the initial section of this chapter, which involves collecting data and ensuring an adequate amount for training, validation, and testing phases. Approximately 220 000 single-particle events are used for training for each particle type, totaling around 880 000 samples, which accounts for 60 % of the complete dataset. The validation and test sets consist of approximately 73 000 single-particle events for each particle type, amounting to a total of 300 000 samples. For this

study, two separate CNN models are trained, one for positive charged tracks and another for negative charged tracks. These CNN models share the same architecture, with the only distinction being the inputs. This distinction is necessary due to the geometry of the ECL, specifically the arrangement of crystals. Positive and negative charged tracks typically traverse a different number of crystals. If all tracks, regardless of their charge, are fed into the CNN, it cannot accurately capture the features specific to each type. Hence, 440 000, 145 000, and 145 000 single-particle events are assigned as the training, validation, and test datasets, respectively, for each CNN model trained on positive and negative charged tracks. To avoid biases toward specific samples in the training, validation, and test dataset, equal numbers of muons and pions are included. This ensures that the model does not favor a particular type of charged track. To insure that samples in each dataset have full coverage of p_T , θ_{ID} , and ϕ_{ID} , distributions of these inputs for samples with $BG \times 1$ for training, validation, and test datasets are shown in Figure 5.4, 5.5, and 5.6, respectively.

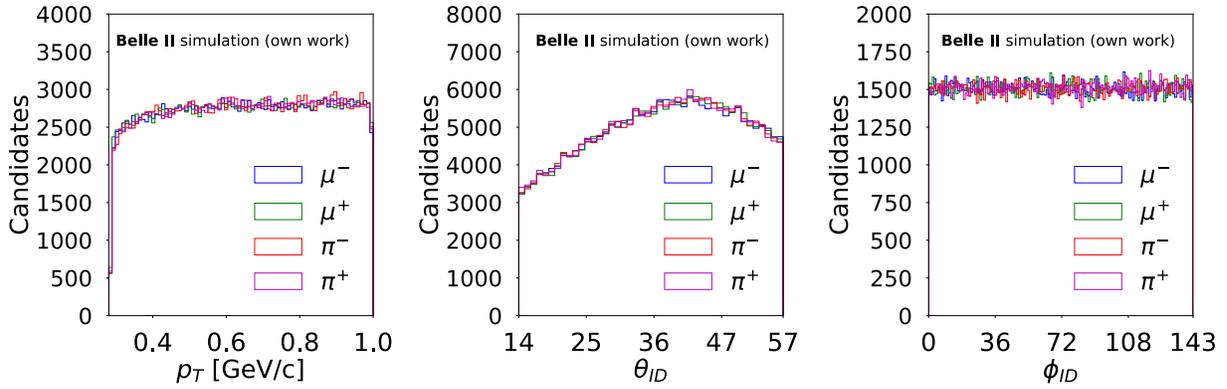


Figure 5.4: Distributions of reconstructed p_T , θ_{ID} , and ϕ_{ID} in training dataset for events generated under $BG \times 1$

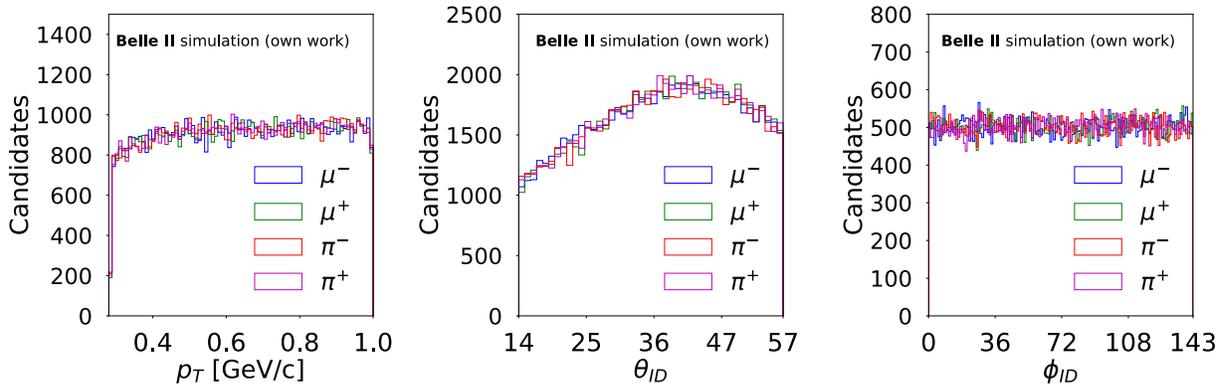


Figure 5.5: Distributions of reconstructed p_T , θ_{ID} , and ϕ_{ID} in validation dataset for events generated under $BG \times 1$

The second part of the preprocessing involves labeling the data. Since supervised learning is utilized in this study, labeling is crucial. In other words, we have knowledge of the inputs and outputs and can assign labels to each type of particle based on the charge

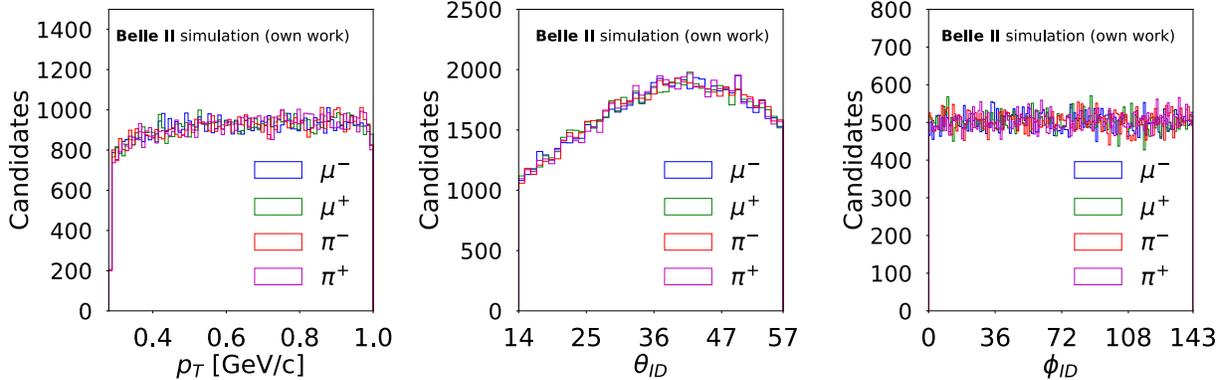


Figure 5.6: Distributions of reconstructed p_T , θ_{ID} , and ϕ_{ID} in test dataset for events generated under $BG \times 1$

and MC information obtained from the generation stage. In this case, muons are assigned the label 1, while pions are labeled as 0. This labeling scheme is general.

The third part of the preprocessing involves identifying and addressing any corrupted or mislabeled images. To accomplish this, pixels with NaN (Not a Number) values are replaced with zero, which is physically accurate since no energy is deposited in these pixels. All input data is examined, and no missing data is found in any of the samples.

The fourth part of the preprocessing involves normalization, which is crucial for neural networks. Neural networks tend to perform better when the input data is within the range of 0 to 1. The energy values in the pixel images are used as they are, without any applied scaling, as they already have small values. However, some pixel images of pions contain very large energy values exceeding 1 GeV due to pion inelastic interactions with nuclei, which produce protons. These large values are replaced with 1 GeV. Since these pixels are very few in number, this adjustment is negligible, affecting the standard deviation and mean of the energy depositions by only 0.09% and 0.005%, respectively. A threshold value of 1 MeV is applied to the energy depositions in the pixels, so that any pixel below this threshold is assigned zero energy. This is to avoid noises that arise due to the high beam background within the Belle II experiment.

Table 5.2 presents the mean and standard deviation values for the training dataset, reflecting the percentage differences in deposited energy for μ^\pm and π^\pm particles between original and modified values. The mean values signify the average alteration in energy deposition after specific replacements. For instance, for μ^- particles within a 7×7 pixel size, the average change in deposited energy is approximately 0.077% post-replacement. Standard deviation values offer a glimpse into the variability of these changes across the dataset. A heightened standard deviation indicates a wider variability in the changes. For π^- particles with a 7×7 pixel size, the standard deviation of the changes is roughly 0.131%, showing some variability in the effects of value replacements. However, these figures are quite minimal and can be disregarded as the changes are on the order of keV.

In addition to the aforementioned steps, there exist other critical preprocessing procedures involving image resizing and data augmentation. The necessity of resizing images

Table 5.2: Mean and standard deviation of percentage differences in deposited energy for μ^\pm and π^\pm particles between original and modified values.

		μ^-	π^-	μ^+	π^+
$\frac{ \Delta E }{E}$ (%)	mean	0.077	0.068	0.078	0.067
	standard deviation	0.008	0.131	0.005	0.266

arises when the images are not of uniform dimensions. However, this aspect is not applicable to the current study since all the images possess a consistent size of 7×7 . On the other hand, data augmentation, a technique used to enhance model performance by artificially expanding the training dataset, does not find relevance here. Normally, this augmentation includes operations such as rotation, zooming, shifting, and flipping images. Yet, given the unique characteristics of the data in this study, caution is exercised. Applying rotations, flips, or shifts to the images could yield unintended outcomes, as maintaining the inherent orientation of the images is paramount.

The p_T values are already constrained within the range of 0.28 to 1.0 GeV/ c . With these values being readily interpretable by neural networks, additional scaling is deemed unnecessary. Another notable point is the handling of variables like θ_{ID} and ϕ_{ID} . Instead of treating them as mere numerical entries, they are considered as categorical entities and are incorporated into the network using embeddings. This choice ensures that the network can recognize and effectively harness their categorical essence, enriching the overall data interpretation.

5.4 Training the CNN

Having established the foundation with input handling and preprocessing, the focus now shifts to the actual training of the neural networks. This phase demands a well-thought-out setup, encompassing the choice of specific hyperparameters. The importance of this choice cannot be understated, as hyperparameters like learning rate, batch size, and architectural configurations significantly dictate the learning dynamics and, by extension, the model's final performance.

5.4.1 Hyperparameter Optimization

The process initiates with the extraction of specific subsets from the main datasets, designated for distinct purposes such as training and validation. A critical step in this procedure is the selection of 20 000 samples for hyperparameter optimization. This selection is conducted through random shuffling, ensuring that the subset is not only representative of the entire dataset but also balanced between muons and charged pions. This approach is particularly important given the time-intensive and resource-heavy nature of hyperparameter optimization. Additionally, a separate subset comprising 4 000 samples is allocated for validation. This validation subset is carefully balanced to include an equal representation

of muons and charged pions, thus ensuring a comprehensive and unbiased assessment of the model’s performance.

Before commencing the training process, the model’s hyperparameters are meticulously fine-tuned using the Optuna hyperparameter optimization library [84]. This optimization routine plays a crucial role in enhancing the model’s performance. Optuna embarks on a search to find the optimal configuration that can minimize the validation loss, which in turn, might enhance the model’s ability to generalize. The optimization approach taken by Optuna hinges on the Tree-structured Parzen Estimator (TPE) algorithm—a Bayesian optimization method. Unlike conventional methods like grid search or random search, TPE dynamically explores the hyperparameter space by drawing insights from previous evaluations. It builds a probabilistic model correlating hyperparameters to the likelihood of achieving a particular score on the objective function, thus recommending new hyperparameter sets based on this model.

In the hyperparameter optimization process, various values are explored for each parameter to fine-tune the model’s performance. Specifically:

- The learning rate (lr), explored within a logarithmic space between 0.00001 and 0.01, is crucial for controlling the step size during the optimization process, affecting the convergence and performance of the trained model.
- The batch size, tested among categorical values of 256, 512, and 1024, influences the gradient estimation, affecting both the convergence speed and the stability of the training process.
- The patience, varied within an integer range of 5 to 30 in steps of 5, governs the adaptation of the learning rate, aiding in avoiding premature stopping or overfitting.
- The initial filters, explored within an integer range of [8, 128], dictate the number of filters in the first convolutional layer, impacting the model’s complexity and capacity to capture features from the input data.
- The embedding dimensions of θ_{ID} and ϕ_{ID} , varied within an integer range of [8, 32], specify the dimensions of the embeddings, influencing the representation capacity of categorical or angular data within the model.
- The number of neurons in fully connected layer 2 (FC2), tested within an integer range of [32, 512], affects the model’s capacity to learn complex representations.
- The dropout rate before FC2, examined among a set of predefined values of 0.05, 0.1, 0.2, and 0.5, controls the dropout regularization, aiding in reducing overfitting by preventing co-adaptation of neurons during training.

Following the optimization process, it is crucial to analyze the progression of the optimization and understand the impact of different hyperparameters on the model’s performance. Figure 5.7 provides a graphical representation of this evolution. The objective

value plotted in this graph is a key performance metric that evaluates how well the model performs under various hyperparameter settings. Each point on the graph represents the model's performance in a specific trial, with successive points indicating the progression of the optimization process. The plot is particularly useful for visualizing the trajectory of the optimization, showing how the model's performance improves or fluctuates with each set of hyperparameters tested. Observing the trend of the objective value helps in understanding the effectiveness of the optimization algorithm in identifying the most suitable hyperparameters. It also indicates how quickly the process converges towards the optimal configuration, providing insights into the stability and efficiency of the hyperparameter tuning strategy.

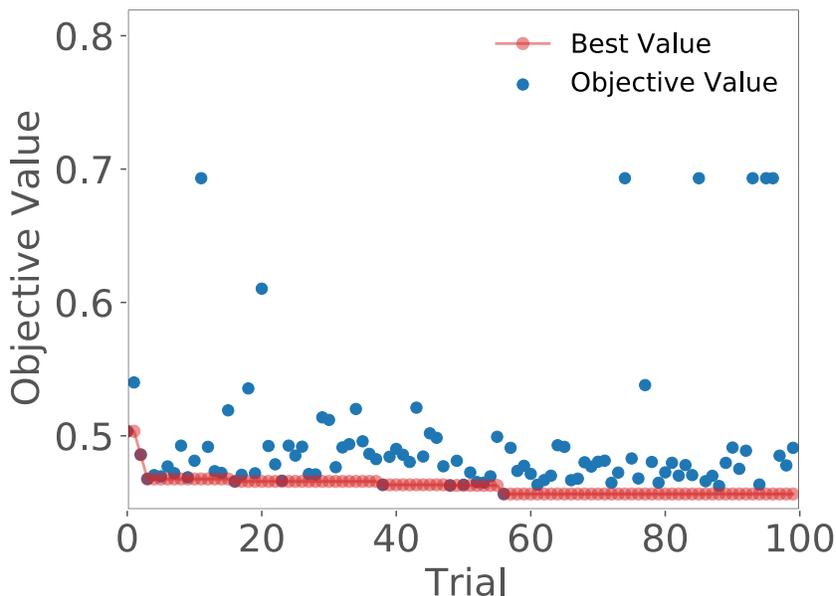


Figure 5.7: Evolution of model's objective value during hyperparameter tuning. This graph illustrates the changes in the model's performance metric, termed as the objective value, through successive trials in the hyperparameter optimization process.

Figure 5.8 elucidates the relative significance of various hyperparameters in attaining optimal model performance. Notably, the initial filters parameter emerges with the highest importance value among others, indicating its substantial impact on the optimization objective. The significance of a hyperparameter is gauged by analyzing the alterations in the objective function values (e.g. validation loss) as the hyperparameter varies across different optimization trials. This analysis underscores the influential role that the initial filters parameter plays in driving the optimization process towards a more favorable outcome, thereby hinting at its potential as a focal point in future tuning exercises.

Finally the best combination of hyperparameters is identified from 100 different combinations. The optimal values are presented in Table 5.3.

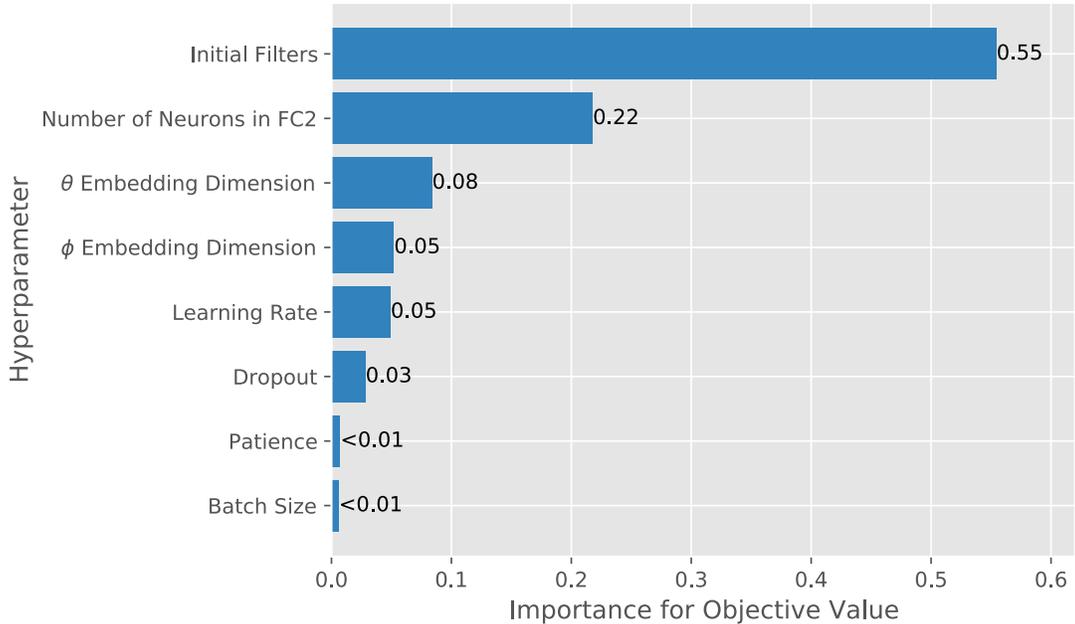


Figure 5.8: This figure presents analysis of the significance of various hyperparameters in optimizing model performance. The chart ranks hyperparameters based on their impact, offering insights into which parameters are most influential in enhancing the accuracy and efficiency of the model. This visualization aids in understanding the model’s sensitivity to specific hyperparameters, facilitating more effective and targeted tuning strategies.

Table 5.3: Optimal hyperparameter values identified through the optimization process.

Hyperparameter	Best value
Batch size	512
θ_{ID} embedding dimension	23
ϕ_{ID} embedding dimension	17
Dropout rate	0.05
Initial filters	117
Learning rate (lr)	0.00145
Number of Neurons in FC2	185
Patience	30

5.4.2 Architecture and Training

Following the optimization of hyperparameters, a robust framework for model training, utilizing PyTorch [85] for both training and testing of the CNN, has been meticulously established, laying a solid foundation for the ensuing training procedure. This framework, empowered by PyTorch’s versatile and efficient computational abilities, is further refined after a comprehensive data cleansing and preparation process. To ensure an equitable representation of the underlying data distribution, 60 % of the dataset, uniformly distributed across both particle types, is allocated for training. Additionally, to provide a robust

platform for continuous performance assessment, 20% of the total dataset is reserved as a validation set. This validation set, balanced in terms of particle types and handled effectively through PyTorch’s data management capabilities, plays an instrumental role in diagnosing any latent issues of overfitting or underfitting, thereby ensuring the reliability and efficacy of the model training framework.

The training process is divided into two parts: convolutional layer and feed-forward neural network (FNN).

In the first stage, 7×7 pixel images of muons and charged pions are used. During convolution, a 3×3 pixel window, known as a kernel, moved across each image. Padding of one pixel is added to the image edges, and the kernel moves with a stride of $(1, 1)$ across the image. The padding adds one pixel with value zero on the edge of the image which is beneficial to capture more information on the edges. After evaluating different counts, 64 initial filters are chosen based on their performance over 100 training epochs.

In the second stage after convolution, the data is flattened and then combined with other parameters such as p_T , θ_{ID} , and ϕ_{ID} . The first layer of the FNN has 3295 neurons, and the second one has 185. Between these layers, a dropout rate of 0.05 is applied to prevent overfitting. Dropout is a very useful tool to excel the performance of the network. The technique is based on removing neurons randomly from a layer. This proves to be effective in this method.

The Adam optimizer with a learning rate of 0.00145 is used. If the validation loss does not improve over 30 consecutive epochs, the learning rate is reduced by half. Cross entropy loss is chosen for this binary classification task. The training is performed for a maximum 100 epochs, with each epoch processing data in batches of 512. The model is saved at the point where it achieves the lowest validation loss during training.

By analyzing the loss figures, it is observed that during the first few epochs, both training and validation losses decreased. However, in the later epochs, while the training loss continued to drop, the validation loss plateaued or even increased, suggesting overfitting. Therefore, for the testing phase, the model from the epoch with the lowest validation loss is selected and saved. A schematic of the network is shown in Figure [5.9](#).

The training and validation accuracy and loss for training using both charged tracks with $BG \times 1$ are shown in Figure [5.10](#). The model in each training is selected based on the lowest validation loss value achieved during 100 epochs of training. There are steps that can be seen in the loss figures which originates from dividing the learning rate by 2 if validation loss value does not improve after 30 epochs of training.

Upon a detailed examination of the loss plots, several observations can be made. During the initial epochs of training, there’s a consistent decrease in both the training and validation loss, signifying that the model is effectively learning from both the training and validation datasets.

However, as training progresses, a divergence in the two losses becomes evident. While the training loss continues to diminish, the validation loss either stagnates or begins to rise. Such behavior is typically indicative of overfitting. At this stage, the model appears

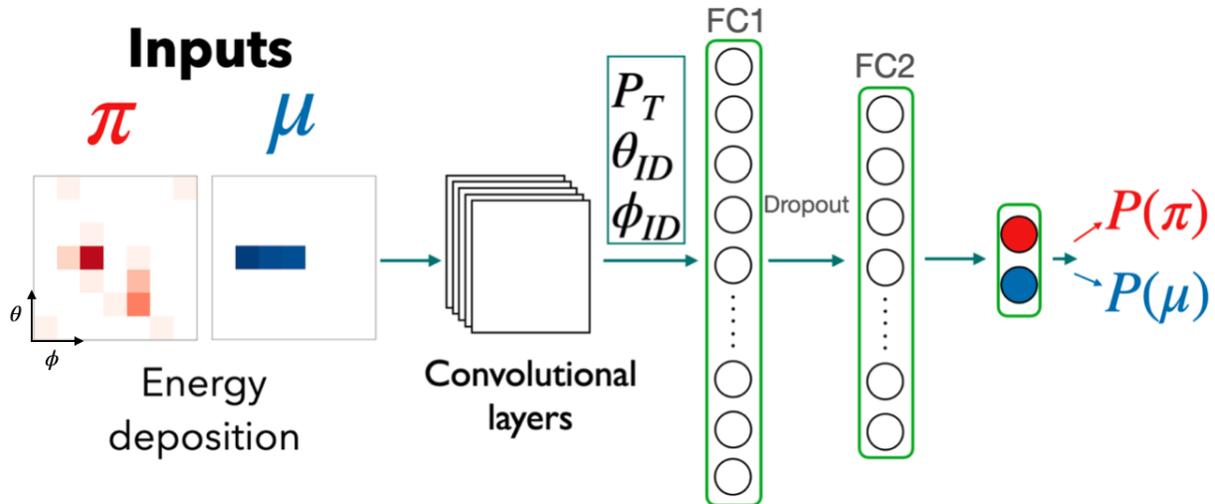


Figure 5.9: CNN architecture. See text for details.

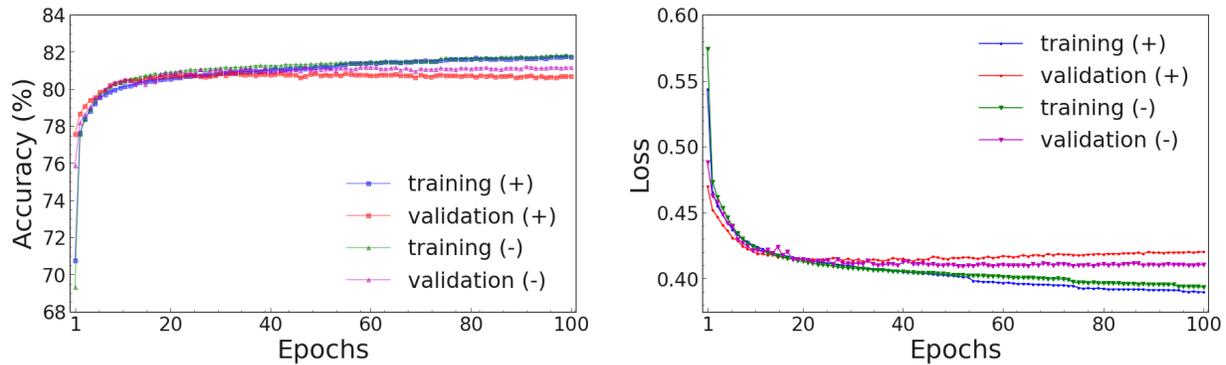


Figure 5.10: Training and validation accuracy (left) and loss (right) for $BG \times 1$.

to be adapting to the specific noise and outliers in the training data, rather than grasping overarching, generalizable patterns.

For the testing phase, it's critical to select an optimal model. Instead of simply choosing the final model after a set number of epochs, the model with the lowest validation loss observed throughout the 100 epochs of training is selected. This approach ensures that the most generalizable version of the model is used for testing.

A salient feature often observable in the loss plots is what's commonly referred to as the inflection point. This is the juncture where the validation loss plateaus or even begins to increase, while the training loss continues to diminish. The appearance of this turning point is a clear sign that overfitting is beginning to take place, providing a concrete indication.

5.5 Model Evaluation

While training metrics and loss plots offer insights, the true measure of a model's performance is its ability to generalize to unseen data. It's not merely about how accurately the

model predicts its training data but how it performs on data it hasn't been exposed to. This emphasizes the importance of dividing dataset into separate training, validation, and testing datasets.

After training the CNNs and choosing the best model based on minimal validation loss, the testing phase becomes paramount. Two primary reasons underscore its significance. First, the testing dataset contains data points not seen during training, ensuring an unbiased evaluation of the model's generalization capability. Second, although the testing phase uses MC samples, which are generated under specific conditions (single track per event) and might not directly mirror real-world unpredictability, they still represent a controlled evaluation setup distinct from the training phase.

When transitioning to the testing phase, it's vital to ensure that the testing dataset undergoes the same preprocessing as the training data to maintain consistency. Once prepared, the testing dataset is introduced to the trained CNN model. In this phase, there's no tweaking of weights; the model merely evaluates the outputs.

Receiver Operating Characteristic

The Receiver Operating Characteristic (ROC) curve is a crucial tool for understanding the performance of classification models, especially for binary classification tasks. The ROC curve is a graphical representation that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) across different threshold settings. The formulae for TPR and FPR are given by:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (5.1)$$

and

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}, \quad (5.2)$$

where:

- TP: is the number of true positives,
- TN: is the number of true negatives,
- FP: is the number of false positives,
- FN: is the number of false negatives.

One of the primary benefits of the ROC curve is that it allows for evaluating a model's performance without committing to a specific threshold. In many real-world scenarios, the optimal threshold may vary based on the application.

Area Under the Curve

The Area Under the Curve (AUC) serves as a concise summary of a model’s ability to distinguish between positive and negative classes, with a perfect classifier achieving an AUC of 1, while a random classifier scores 0.5. In an ideal classification scenario, the classifier attains a true positive rate of 1 without any false positives, corresponding to a random classifier with an AUC of 0.5. The ROC curve reflects a trade-off in balancing true positives and false positives, with the optimal point on the curve depending on the specific problem domain and the costs associated with false positives versus false negatives [86]. The AUC offers a comprehensive view of the model’s discriminative power by considering its ability to rank instances correctly across the entire range of possible thresholds. Unlike some other metrics, such as accuracy, precision, and recall which focus on specific aspects of performance, the AUC provides a holistic assessment. It takes into account the model’s performance at various decision thresholds, offering valuable insights into its overall discriminatory capability. A higher AUC indicates that the model can effectively differentiate between positive and negative instances, making it a crucial metric for evaluating classification performance.

To provide a more granular analysis at critical regions of the ROC curve, a binning strategy is employed. Specifically, 100 working points are designated between the values of 0 and 0.05, and between 0.95 and 1.0, to finely examine the model performance near the extremes. In the range of 0.05 to 0.95, 80 working points are utilized to allow for a thorough yet balanced examination of the model’s discriminative ability across varying thresholds. This custom spacing is facilitated by a dedicated function, ensuring a consistent and precise evaluation across different ROC curves generated in the study. This is the case everywhere throughout the thesis unless stated otherwise.

5.5.1 Muon Efficiency vs. Pion Fake Rate

The CNN performance is assessed using a test dataset containing an equal distribution of muons and charged pions. This dataset is generated and reconstructed under identical conditions to the training and validation datasets.

The ROC curves in Figure 5.11 exhibit the trade-off between muon identification efficiency and the pion fake rate. The formulae for muon efficiency and pion fake rate at a given threshold are given by:

$$\text{Muon efficiency} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{correctly identified muons at a given threshold}}{\text{all muons}}, \quad (5.3)$$

and

$$\text{Pion fake rate} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{\text{pions incorrectly identified as muons at a given threshold}}{\text{all pions}}, \quad (5.4)$$

where:

- TP is the number of correctly identified muons,
- TN is the number of correctly identified pions,
- FP is the number of pions incorrectly identified as muons,
- FN is the number of muons incorrectly identified as pions.

Distinctly, three curves are highlighted:

- All tracks: Encompassing all tracks, regardless of their cluster association.
- Tracks with cluster: Denoting tracks successfully associated with an ECL cluster.
- Tracks without cluster: Indicating tracks without a corresponding ECL cluster, relying purely on extrapolation.

Particularly in the case of "Tracks without cluster" where challenges in extrapolation are most evident, the complexities of track cluster matching within the ECL become pronounced. This stage, involving extrapolation and preliminary matching to the ECL's calibrated data points, is susceptible to inaccuracies due to the intricate interplay of track curvature and the specific geometry of the calorimeter's crystals. This complexity can result in multiple or no associations with the calorimeter's data points. Moreover, the process must accurately account for various particle interaction hypotheses within the connected regions of the ECL, ranging from single ionizing particles with additional photons to interactions involving charged hadrons or electron/positron combinations. The culmination of these challenges is most noticeable in the final matching phase, where the goal is to accurately align reconstructed tracks with corresponding shower in the ECL. This phase is particularly demanding, especially when multiple showers may be linked to a single track, requiring sophisticated analytical methods for correct interpretation and encoding of this multifaceted information. These factors collectively highlight the challenging nature of track cluster matching in the ECL.

Notably, the "Tracks without cluster" curve achieves superior muon efficiency at lower pion fake rates compared to its counterparts. This observation underscores the CNN's proficiency in discerning muons from charged pions when tracks lack clusters. Although the distinction between "All tracks" and "Tracks with cluster" is subtle, it emphasizes the critical role of cluster presence in the model's discrimination potential.

In Tables [5.4](#) and [5.5](#), muon efficiencies at fixed pion fake rates and pion fake rates at fixed muon efficiencies are presented for positive and negative charged tracks, respectively. These values are derived from three track categories: "All tracks", "Tracks with cluster", and "Tracks without cluster". Each category encapsulates a distinct aspect of the CNN model's performance across different charge types. Further visual insight is provided by Figure [5.12](#), which offers a zoomed-in visualization of the ROC curve, emphasizing regions

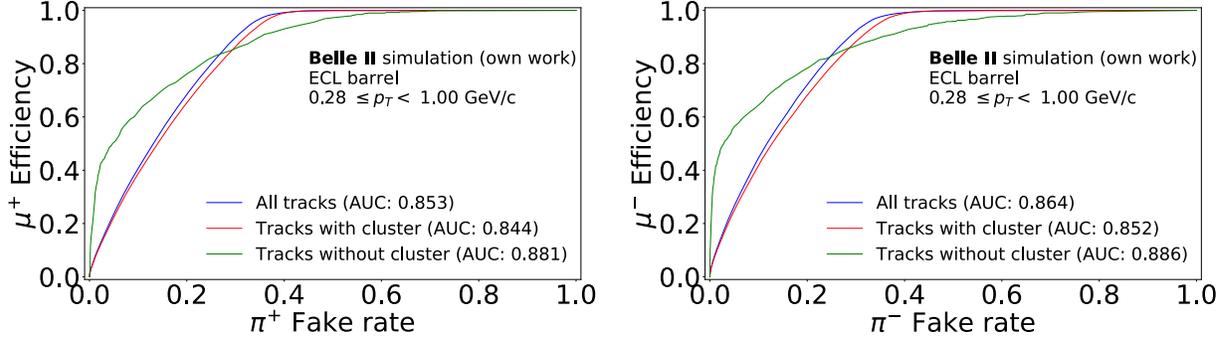


Figure 5.11: ROC curve illustrating the performance of the CNN model in distinguishing muons from charged pions using 7×7 pixel energy deposition images in the ECL. The plot encompasses tracks of both positive (left) and negative charges (right).

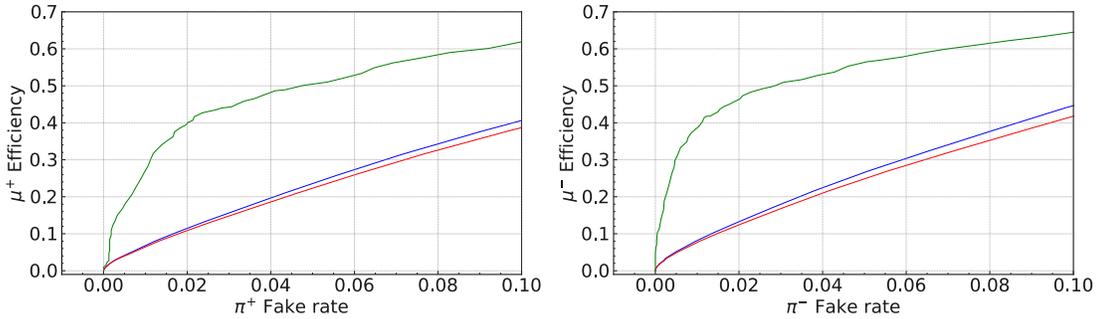


Figure 5.12: Zoomed-in view of the ROC curve from the above figure, specifically highlighting the region of low pion fake rates, which is crucial for precision measurements and rare decay searches.

with low pion fake rates and underscoring the model’s discriminating capacity in these critical areas.

The stark behavior of the green curve indicates the presence of invaluable information within tracks devoid of matched clusters. Potential reasons include inefficiencies in clustering, cluster-track matching challenges, or certain tracks genuinely lacking clusters.

The discrepancy in muon efficiencies between positive and negative tracks, particularly at small pion fake rates, highlights the nuanced particle interactions within the ECL. A significant determinant for these disparities is the distinct trajectories charged particles adopt within the detector, governed by its magnetic field and geometry. This divergence causes differently charged particles to traverse varying numbers of ECL crystals, culminating in unique energy deposition patterns. Such patterns, due to their distinctiveness, offer the CNN model valuable discriminative features.

5.5.2 Evaluation in Different Transverse Momentum Ranges

To gain a deeper insight into the intricacies of PID in various momentum regimes, the performance analysis are divided into distinct p_T regions.

Table 5.4: μ^+ efficiencies at fixed π^+ fake rates (top) and π^+ fake rates at fixed μ^+ efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/c for all tracks with CNN output, tracks with cluster, and tracks without cluster. Fix values 1 %, 5 %, 10 %, and 20 % are selected for π^+ fake rates and 80 %, 90 %, 95 %, and 99 % for μ^+ efficiencies.

Category	μ^+ efficiency at fixed π^+ fake rates (%)			
	1 %	5 %	10 %	20 %
All tracks	4.99	20.06	35.02	61.29
Tracks with cluster	4.81	19.38	33.91	59.40
Tracks without cluster	4.76	49.86	61.90	80.95

Category	π^+ fake rate at fixed μ^+ efficiencies (%)			
	80 %	90 %	95 %	99 %
All tracks	28.56	33.93	37.03	40.87
Tracks with cluster	29.71	35.28	38.51	42.30
Tracks without cluster	21.56	31.33	38.09	39.85

Low p_T Region ($0.28 \leq p_T < 0.52$ GeV/c):

In this region, particles exhibit pronounced curved trajectories due to the magnetic field's influence, leading to unique energy deposition patterns in the ECL. Tracks in this momentum range are less likely to produce distinct clusters that can be matched, given their lower energy and more curved trajectories. Hence, the model's performance might be more influenced by the energy deposition patterns rather than matched clusters. ROC curves corresponding to this region is shown in Figure [5.13](#).

Intermediate p_T Region ($0.52 \leq p_T < 0.76$ GeV/c):

The trajectories of particles in this momentum range are somewhat curved but less so compared to the lower p_T region. As the momentum increases, the chances of forming and matching clusters increase, enhancing the discrimination between muons and charged pions. However, the CNN model still plays a crucial role, especially for tracks without matched clusters. ROC curves corresponding to this region is shown in Figure [5.14](#).

Table 5.5: μ^- efficiencies at fixed π^- fake rates (top) and π^- fake rates at fixed μ^- efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/c for all tracks with CNN output, tracks with cluster, and tracks without cluster. Fix values 1 %, 5 %, 10 %, and 20 % are selected for π^- fake rates and 80 %, 90 %, 95 %, and 99 % for μ^- efficiencies.

Category	μ^- efficiency at fixed π^- fake rates (%)			
	1 %	5 %	10 %	20 %
All tracks	5.10	20.96	36.21	63.66
Tracks with cluster	4.90	20.22	34.93	61.48
Tracks without cluster	0.00	23.53	35.29	63.62

Category	π^- fake rate at fixed μ^- efficiencies (%)			
	80 %	90 %	95 %	99 %
All tracks	27.45	32.87	36.03	40.47
Tracks with cluster	28.72	34.39	37.68	42.25
Tracks without cluster	42.74	57.31	78.66	95.73

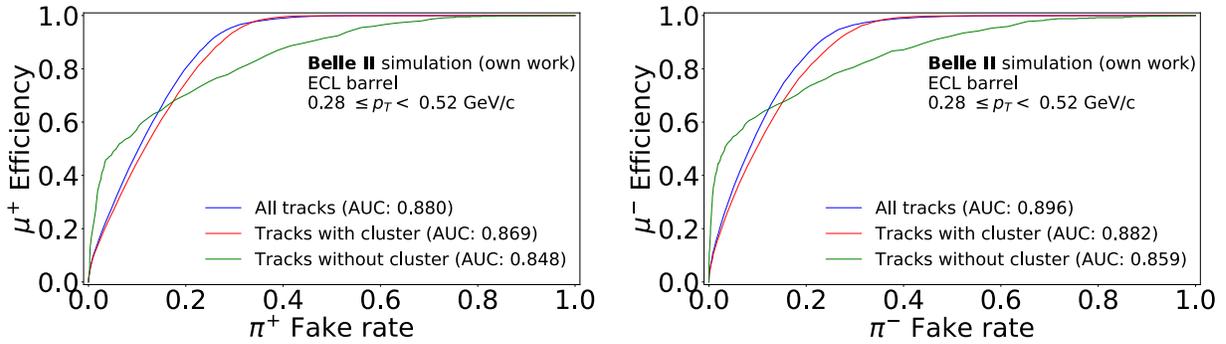


Figure 5.13: ROC curve illustrating the performance of the CNN model in distinguishing muons from charged pions in low p_T region. The plot encompasses tracks of both positive (left) and negative charges (right).

High p_T Region ($0.76 \leq p_T < 1.0$ GeV/c):

Particles in this region can partly reach the KLM, responsible for muon identification. This can influence the energy deposition patterns seen in the ECL. Moreover, tracks with higher p_T have a greater likelihood of being matched to clusters. Given the higher likelihood of tracks being matched to clusters, there's a reduced sample of tracks without matched clusters. This limited statistics might lead to increased uncertainties and potentially lower

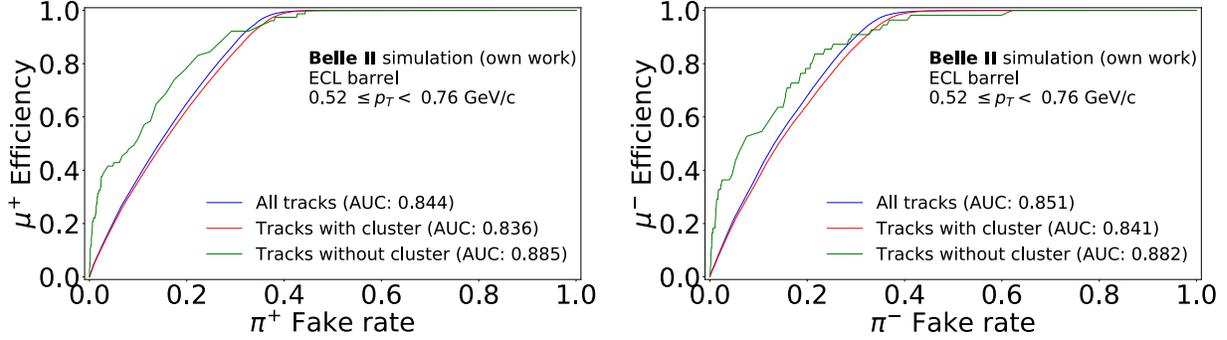


Figure 5.14: ROC curve illustrating the performance of the CNN model in distinguishing muons from charged pions in intermediate p_T region. The plot encompasses tracks of both positive (left) and negative charges (right).

AUC for the green curve (representing "Tracks without cluster"). The interactions with the KLM can also introduce additional complexities in energy deposition patterns, influencing the model's performance. ROC curves corresponding to this region is shown in Figure 5.15.

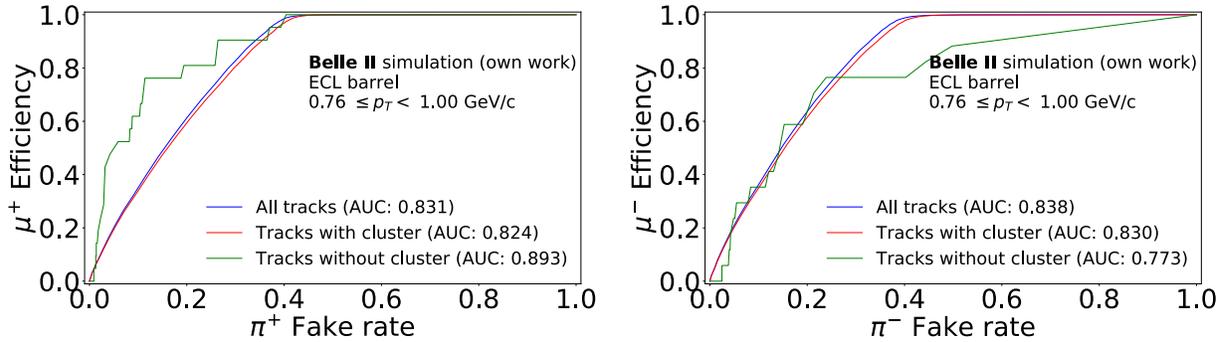


Figure 5.15: ROC curve illustrating the performance of the CNN model in distinguishing muons from charged pions in high p_T region. The plot encompasses tracks of both positive (left) and negative charges (right).

For all regions, considering the correlation between p_T and the likelihood of having matched clusters is crucial. The increasing probability of cluster matching as p_T increases means that the model's reliance on purely energy deposition patterns decreases in higher p_T regions. However, for tracks without matched clusters, the model's performance becomes even more crucial, especially in the high p_T region where statistics are limited.

In summary, the interplay between p_T , cluster matching likelihood, and the model's performance provides a comprehensive view of the challenges and intricacies in μ/π separation across different momentum scenarios.

5.6 Comparison with Other PID Methods

PID is a critical component in the data analysis pipeline of the Belle II software. In this context, two primary PID methods are employed: the standard binary PID and the binary BDT PID. The standard binary PID utilizes the likelihoods and the BDT PID method relies on clustering information and shower shape variables derived from the ECL. For clarity, the binary BDT PID method will henceforth be referred to simply as "BDT" and the standard binary PID method as default. The BDT method capitalizes on the shower shape variables of particles interacting with the ECL, making the presence of a distinct cluster crucial for accurate predictions. In contrast, the CNN PID method represents a significant shift from this approach. It does not depend exclusively on cluster data. Instead, by analyzing patterns of energy deposition and other relevant features, the CNN model offers a more comprehensive approach to PID. This is particularly advantageous in situations where tracks cannot be directly linked to specific clusters. To ensure a fair comparison between these methods, it's important to note that both the default and BDT PID methods are applied exclusively within the context of the ECL.

The ROC curves in Figure 5.16, spanning a p_T range of 0.28 to 1 GeV/ c , provide a comparative analysis of the CNN, BDT, and default PID methods for both positive and negative charged tracks. The CNN curve's pronounced advantage accentuates its superior discriminative power, even when only considering tracks with associated clusters.

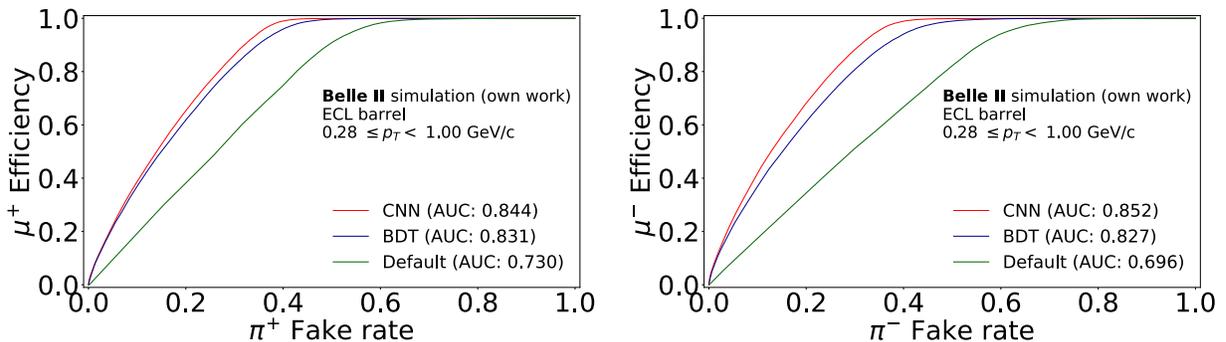


Figure 5.16: ROC curves illustrating the discriminative performance of the CNN PID method compared to the BDT and default PID methods over a p_T range of 0.28 to 1 GeV/ c . The left and right plots correspond to positive and negative charged tracks, respectively, showing the nuanced efficiencies and fake rates across different charge types.

5.6.1 Efficiencies and Fake Rates at Fixed Values

While visual insights from the ROC curves are enlightening, a quantitative analysis offers a deeper understanding of the performance differences.

Performance metrics at fixed values of efficiencies or fake rates elucidate the behavior of the PID models. For example, at a fixed μ^+ efficiency of 90%, the CNN PID method's pion fake rate is 32.03%. In contrast, the BDT and default PID methods register fake

rates of 34.78 % and 49.30 % respectively, underscoring the CNN’s enhanced discriminative capabilities. Tables 5.6 and 5.7 provide a detailed breakdown of efficiencies and fake rates for both charge types at fixed values of efficiencies and fake rates.

Table 5.6: μ^+ efficiencies at fixed π^+ fake rates (top) and π^+ fake rates at fixed μ^+ efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for different PID methods: CNN, BDT, and default. Fix values 1 %, 5 %, 10 %, and 20 % are selected for π^+ fake rates and 80 %, 90 %, 95 %, and 99 % for μ^+ efficiencies.

PID method	μ^+ efficiency at fixed π^+ fake rates (%)			
	1 %	5 %	10 %	20 %
CNN	6.44	22.31	38.69	65.29
BDT	6.12	21.52	36.95	61.78
Default	1.62	9.46	19.26	38.17

PID method	π^+ fake rate at fixed μ^+ efficiencies (%)			
	80 %	90 %	95 %	99 %
CNN	26.75	32.03	35.32	40.06
BDT	28.73	34.78	39.03	45.76
Default	42.75	49.30	54.38	62.94

The CNN’s proficiency in leveraging energy deposition patterns becomes more pronounced when compared with the BDT and default PID methods. Its ability to differentiate between the unique energy deposition patterns of both charges, influenced by the detector’s magnetic fields and geometry, accentuates its adaptability. Such flexibility positions the CNN model as a potent tool, potentially revolutionizing the PID landscape at Belle II.

5.6.2 Evaluation in Different Transverse Momentum Regions

The evaluation spans various momentum ranges and charge types, providing comprehensive insights into the capabilities of the CNN. Three distinct momentum ranges are considered: low, intermediate, and high p_T . The performance is evaluated for both positive and negative charged tracks, with separate CNN models trained for each charge type.

Figures 5.17 depict the ROC curves for negative and positive charged tracks, respectively, across the three p_T ranges.

From the ROC curves, it is evident that the CNN consistently outperforms both the BDT and default PID methods across all momentum ranges and charge types. The dis-

Table 5.7: μ^- efficiencies at fixed π^- fake rates (top) and π^- fake rates at fixed μ^- efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for different PID methods: CNN, BDT, and default. Fix values 1%, 5%, 10%, and 20% are selected for π^- fake rates and 80%, 90%, 95%, and 99% for μ^- efficiencies.

PID method	μ^- efficiency at fixed π^- fake rates (%)			
	1%	5%	10%	20%
CNN	7.66	24.84	41.79	68.16
BDT	6.88	22.20	36.91	61.19
Default	1.72	8.88	17.44	34.44

PID method	π^- fake rate at fixed μ^- efficiencies (%)			
	80%	90%	95%	99%
CNN	25.49	31.10	34.65	40.29
BDT	29.55	36.22	41.05	50.22
Default	48.56	55.72	61.21	71.62

tion is particularly pronounced in the low momentum range, highlighting the CNN’s capability to handle the challenges posed by overlapping particle signatures.

To provide a better understanding of the performance, the AUC metric derived from the ROC curves is utilized. The AUC offers a singular value, facilitating direct comparisons between methods. The percentage improvement of the CNN over other methods (BDT, default) is calculated using the formula:

$$\text{Improvement (\%)} = \left(\frac{\text{AUC}_{\text{CNN}} - \text{AUC}_{\text{Method}}}{\text{AUC}_{\text{Method}}} \right) \times 100, \quad (5.5)$$

where AUC_{CNN} is the AUC value for the CNN and $\text{AUC}_{\text{Method}}$ is the AUC value for the method being compared against (either BDT or default PID).

The percentage improvements of the CNN over the BDT and default PID methods are calculated and shown in Table 5.8. The improvement is more evident in intermediate p_T range for both charged tracks in case of comparing with BDT. In comparison with default method it is more apparent in low p_T range.

The compelling performance of CNNs in PID processes, particularly in contexts akin to Belle II, is evident. Their ability to offer reduced pion fake rates and heightened muon efficiencies suggests potential for diminishing systematic uncertainties, fostering more precise physics analyses.

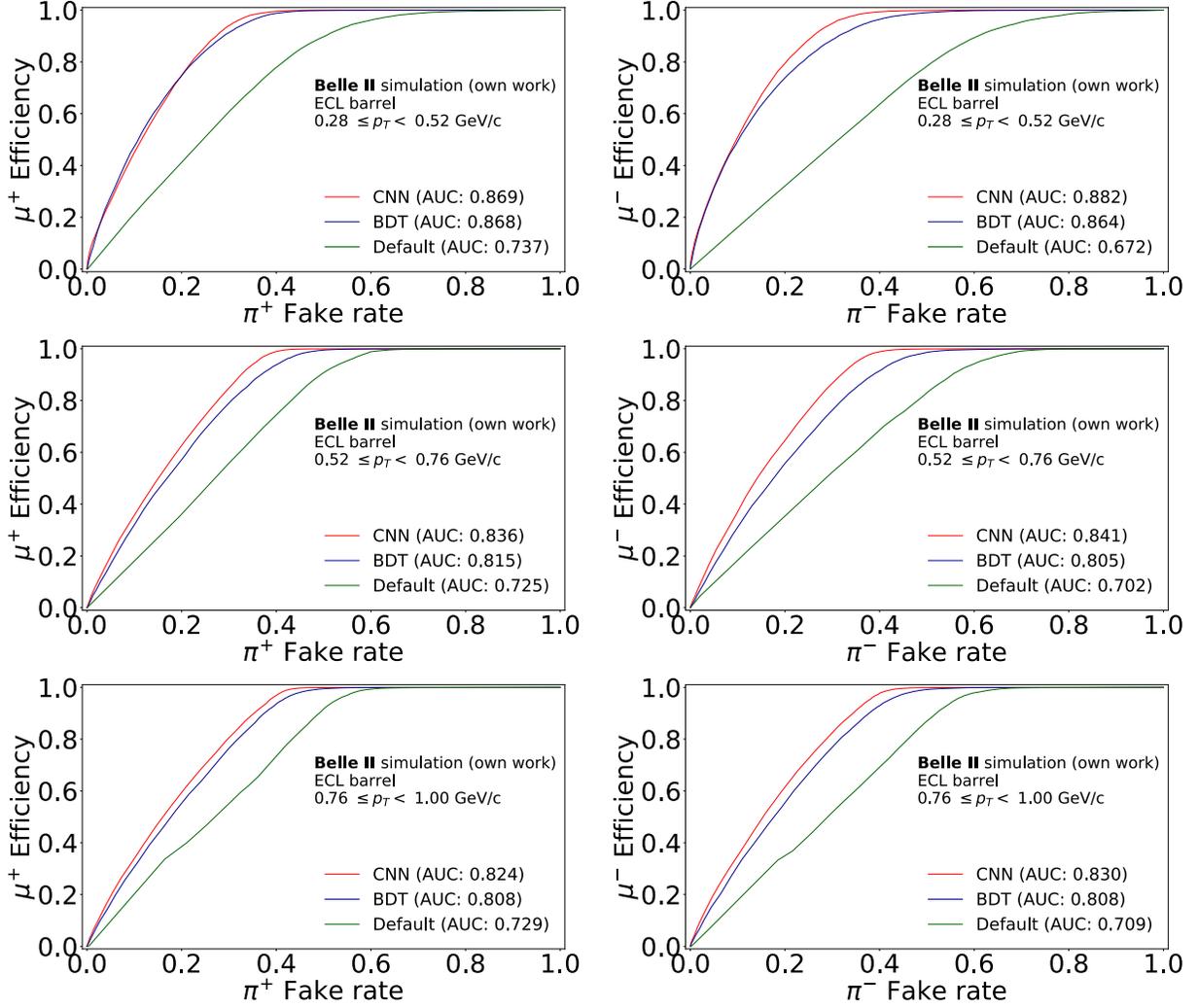


Figure 5.17: ROC curves for positive (left) and negative (right) charged tracks across different p_T ranges for different PID methods. From top to bottom: low ($0.28 \leq p_T < 0.52$ GeV/c), intermediate ($0.52 \leq p_T < 0.76$ GeV/c), and high ($0.76 \leq p_T < 1.0$ GeV/c).

5.6.3 Comparative Analysis of CNN and BDT in the ECL Barrel

To delve deeper into the ECL barrel's different regions, the barrel is divided into three distinct sections: two near the edges and one central region. Distinct combinations of p_T and θ_{ID} ranges are utilized to calculate AUC values for both the CNN and BDT PID methods. This comparative analysis considers only those tracks with corresponding values between the two methods.

For a comprehensive overview, 2D plots of AUC values are presented. Figures 5.18 and 5.19 illustrate the AUC values for positive charged tracks, while Figures 5.20 and 5.21 depict those for negative charged tracks.

The descriptive statistics of the AUC values are summarized in Table 5.9.

Table 5.8: Percentage improvements of the CNN over BDT and default PID methods based on AUC values.

Charge	p_T range	Improvement over BDT (%)	Improvement over default (%)
+	Low	0.15	17.95
	Intermediate	2.60	15.34
	High	2.01	13.04
Charge	p_T range	Improvement over BDT (%)	Improvement over default (%)
-	Low	2.07	31.24
	Intermediate	4.46	19.83
	High	2.77	17.08

Table 5.9: Descriptive statistics for AUC values of CNN and BDT methods in different p_T and θ_{ID} ranges.

Charge	PID Method	Mean	Standard Deviation	Minimum	Maximum
+	CNN	0.843	0.017	0.820	0.878
	BDT	0.830	0.024	0.806	0.876
-	CNN	0.850	0.020	0.823	0.890
	BDT	0.825	0.024	0.797	0.876

The CNN method consistently outperforms the BDT in terms of mean AUC values for both charge polarities. This superior performance underscores the importance of spatial patterns in energy depositions for discriminating between muons and charged pions. CNN’s ability to recognize these patterns even in regions with complex deposition profiles highlights its robustness.

The BDT method exhibits greater variability, possibly due to its reliance on the shower shape variables. In areas where cluster formation is non-trivial, such as near the edges, the BDT’s performance may diminish, reflected in lower AUC values.

Edge effects, particularly around θ_{ID} ’s 14 and 57, pose significant challenges in the ECL barrel due to the barrel’s geometry and the tilted orientation of the crystals. These factors influence particle trajectories and the resulting energy deposition patterns. High-energy particles typically interact with fewer crystals, leading to more concentrated energy depositions, while lower-energy particles may traverse multiple crystals due to the tilt and

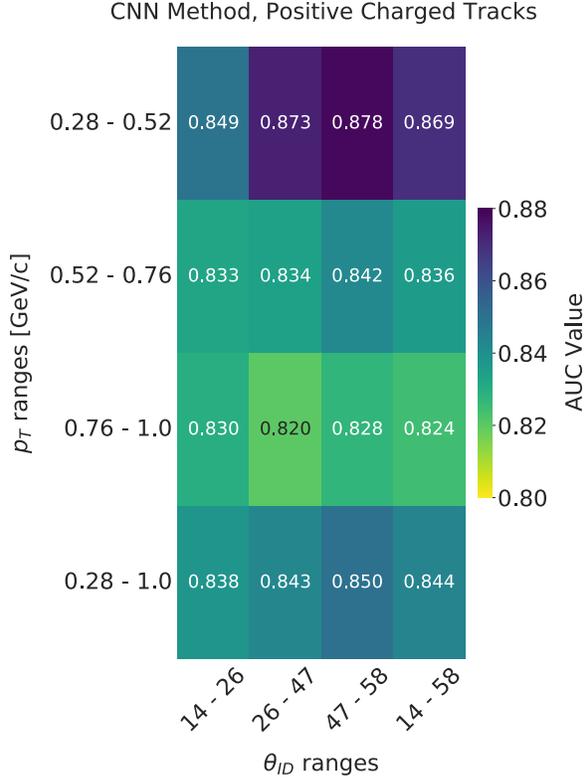


Figure 5.18: AUC values obtained using the CNN PID method for positively charged muons and pions, presented across different p_T and θ_{ID} ranges. Regions with higher intensity colors indicate higher AUC values, reflecting stronger model performance in those areas.

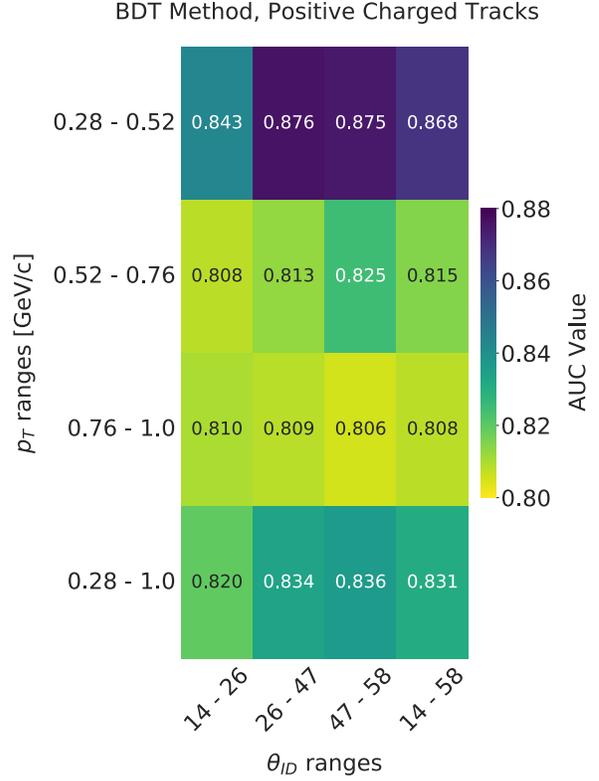


Figure 5.19: AUC values obtained using the BDT PID method for positively charged muons and pions, presented across different p_T and θ_{ID} ranges. Regions with higher intensity colors indicate higher AUC values, reflecting stronger model performance in those areas.

unique geometry, causing energy to be deposited beyond the central 7×7 image area. This results in varied and spread-out energy patterns, complicating the discrimination process.

In summary, this analysis underscores the strengths and limitations of the CNN and BDT PID methods for μ/π separation within the ECL barrel. The interplay between the computational techniques and the ECL barrel's physical attributes provides deep insights into the nuances of particle interactions and the energy patterns they produce.

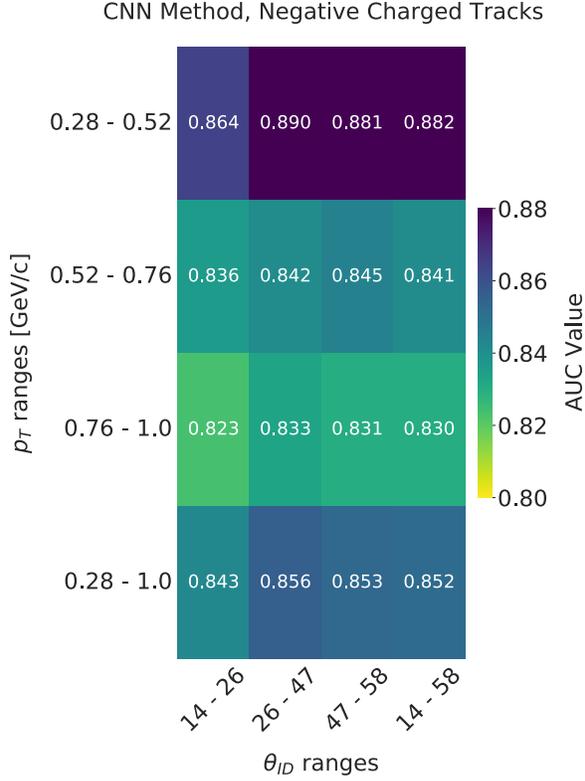


Figure 5.20: AUC values obtained using the CNN PID method for positively charged muons and pions, presented across different p_T and θ_{ID} ranges. Regions with higher intensity colors indicate higher AUC values, reflecting stronger model performance in those areas.

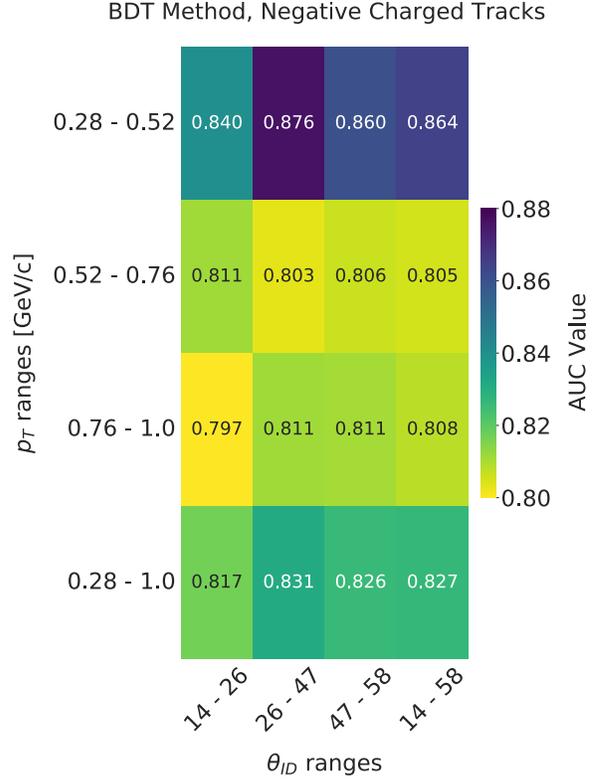


Figure 5.21: AUC values obtained using the BDT PID method for positively charged muons and pions, presented across different p_T and θ_{ID} ranges. Regions with higher intensity colors indicate higher AUC values, reflecting stronger model performance in those areas.

5.7 Energy Thresholds' Effect

In the preprocessing stage of ML models within particle physics, the selection of energy thresholds is crucial, particularly for distinguishing between muons and charged pions. This practice involves balancing the reduction of noise against the retention of vital signals. The choice of threshold is especially pivotal in low momentum regions, as it significantly impacts the model's ability to differentiate between these particles. An important aspect to investigate is how altering the energy threshold during the evaluation phase, while maintaining a consistent training environment, can reveal the model's sensitivity and adaptability.

5.7.1 Significance of Energy Thresholds

The energy threshold is more than a mere filter; it's the prism through which the model interprets the data. The chosen threshold directly affects the model's ability to differentiate

between muons and charged pions, especially in low momentum regions. This distinction becomes pronounced when analyzing the inherent energy distributions of muons and charged pions, as depicted in Figures 5.22 and 5.23. These images, captured at different energy thresholds, elucidate how the model’s interpretative prowess is shaped by the threshold value. As the threshold escalates, images of muons and charged pions become increasingly akin, suggesting a decline in their separation.

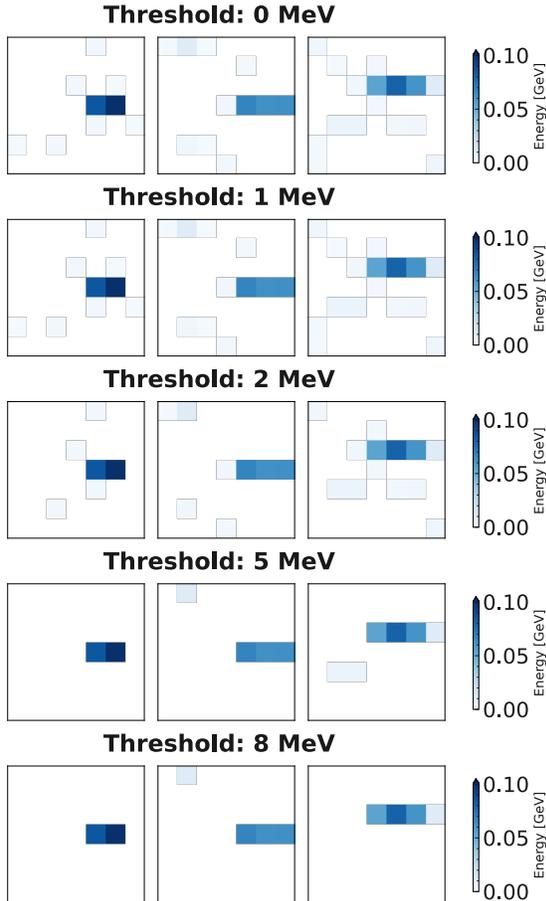


Figure 5.22: 7×7 pixel images of μ^+ at various thresholds: 0, 1, 2, 5, 8 MeV. No threshold is shown as 0 MeV. The threshold is increased from top to bottom.

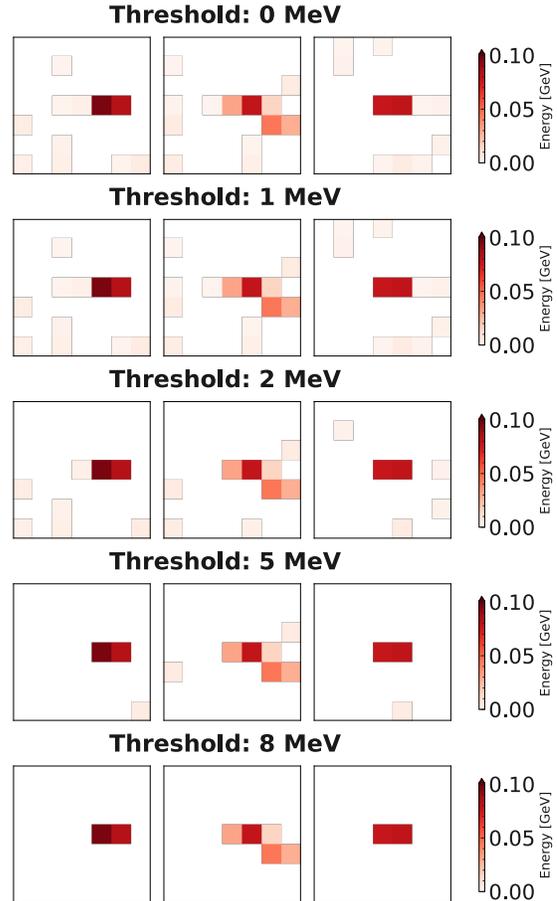


Figure 5.23: 7×7 pixel images of π^+ at various thresholds: 0, 1, 2, 5, 8 MeV. No threshold is shown as 0 MeV. The threshold is increased from top to bottom.

5.7.2 Trade-off Between Efficiency and Fake Rate

The ROC curves in Figure 5.24 highlight the relationship between muon efficiency and pion fake rate as the energy threshold heightens. There’s a discernible dip in efficiency at a constant pion fake rate, emphasizing the model’s sensitivity to the energy threshold. Particularly at higher threshold values like 5 MeV and 8 MeV, a decline in performance is

evident. Noteworthy is the slight superiority in performance at 0 MeV threshold, alluding to the model’s ability to utilize even subtle signals in the data. Further visual insight is provided by Figures 5.25 and 5.26, which offer a zoomed-in visualization of the ROC curve, emphasizing regions with low pion fake rates and underscoring the model’s discriminating capacity in these critical areas.

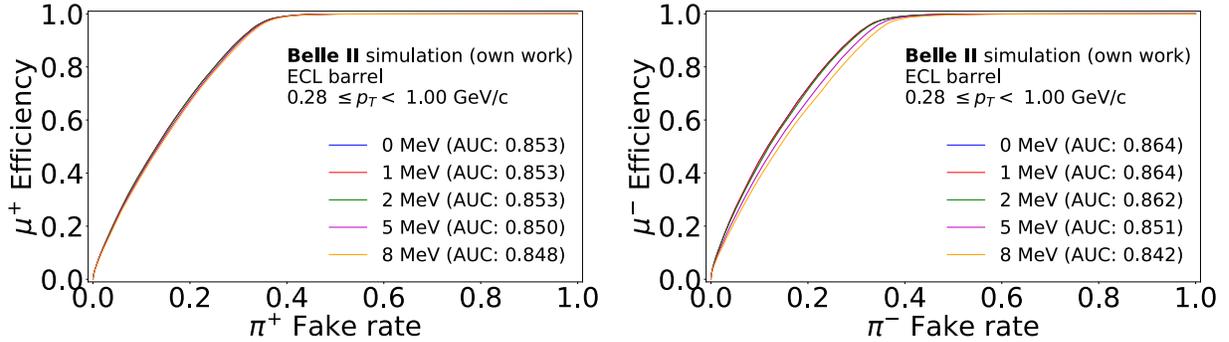


Figure 5.24: ROC curves showing the discriminative performance of the CNN-based PID method across various energy thresholds. Each curve represents a distinct threshold, demonstrating how the model’s sensitivity to muon and charged pion discrimination varies with the energy threshold value.

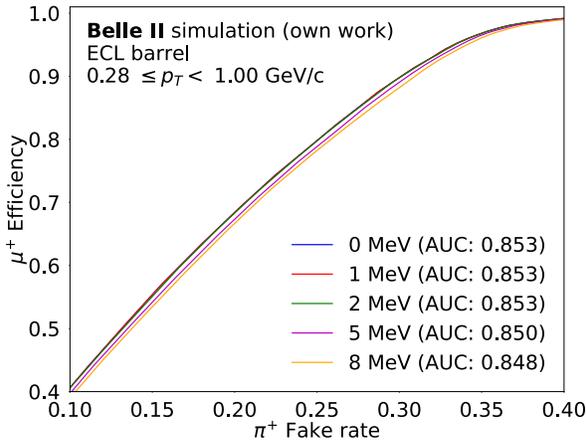


Figure 5.25: Zoomed-in view of the ROC curves for various energy thresholds, specifically highlighting the region of low π^+ fake rates.

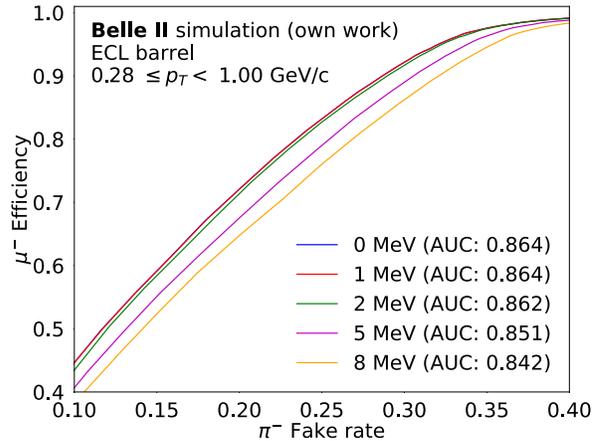


Figure 5.26: Zoomed-in view of the ROC curves for various energy thresholds, specifically highlighting the region of low π^- fake rates.

Through examining the ROC curve values, a discernible pattern surfaces. The efficiency for both positively and negatively charged tracks diminishes as the energy threshold augments. For instance, at a 1% π^+ fake rate, μ^+ efficiency descends from 6.79% to 6.33% as the energy threshold transitions from 0 MeV to 8 MeV. Likewise, when μ efficiency is fixed at 99%, the π fake rate escalates for both charge polarities as the energy threshold increases.

Tables [5.10](#) and [5.11](#), which present the muon efficiency and pion fake rate across various energy thresholds and charge polarities, highlight the complex relationship between the energy threshold and the model's ability to discriminate between particle types.

Table 5.10: μ^+ efficiencies at fixed π^+ fake rates (top) and π^+ fake rates at fixed μ^+ efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for various energy thresholds. Fix values 1 %, 5 %, 10 %, and 20 % are selected for π^+ fake rates and 80 %, 90 %, 95 %, and 99 % for μ^+ efficiencies.

Threshold (MeV)	μ^+ efficiency at fixed π^+ fake rates (%)			
	1 %	5 %	10 %	20 %
0	6.79	23.58	40.63	68.42
1	6.80	23.57	40.63	68.41
2	6.70	23.47	40.59	68.31
5	6.58	22.92	39.69	67.33
8	6.33	22.43	38.99	66.69

Threshold (MeV)	π^+ fake rate at fixed μ^+ efficiencies (%)			
	80 %	90 %	95 %	99 %
0	25.09	30.13	33.31	39.32
1	25.09	30.11	33.32	39.34
2	25.11	30.14	33.37	39.37
5	25.55	30.52	33.72	39.67
8	25.85	30.95	34.11	40.14

5.7.3 Performance Across Transverse Momentum Ranges

The AUC values across different energy thresholds are shown in Figure [5.27](#). Examining AUC values across distinct energy thresholds reveals a slight performance degradation as the threshold amplifies, particularly for negatively charged tracks in lower p_T range (0.28 - 0.52 GeV/ c). This trend persists across intermediate and high p_T ranges, manifesting a gradual AUC decline with rising thresholds. The AUC values are presented in Table [5.12](#).

5.7.4 Implications on Event Complexity

The threshold serves as a signal-to-noise arbitrator. Elevated thresholds, while filtering out noise, might forfeit informative signals, diminishing efficiency. In more complex events

Table 5.11: μ^- efficiencies at fixed π^- fake rates (top) and π^- fake rates at fixed μ^- efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for various energy thresholds. Fix values 1 %, 5 %, 10 %, and 20 % are selected for π^- fake rates and 80 %, 90 %, 95 %, and 99 % for μ^- efficiencies.

Threshold (MeV)	μ^- efficiency at fixed π^- fake rates (%)			
	1 %	5 %	10 %	20 %
0	8.21	26.57	44.63	72.10
1	8.18	26.59	44.64	72.07
2	8.12	25.98	43.46	71.37
5	7.35	23.63	40.68	67.53
8	6.78	21.90	38.15	64.75

Threshold (MeV)	π^- fake rate at fixed μ^- efficiencies (%)			
	80 %	90 %	95 %	99 %
0	23.44	28.66	32.06	39.12
1	23.46	28.68	32.07	39.09
2	23.73	28.96	32.27	39.37
5	25.46	30.55	33.83	41.02
8	26.86	32.08	35.35	43.10

Table 5.12: AUC values for positive and negative charged tracks at different energy thresholds across various p_T ranges.

Charge	p_T range (GeV/ c)	0 MeV	1 MeV	2 MeV	5 MeV	8 MeV
+	0.28 - 0.52	0.8804	0.8804	0.8797	0.8779	0.8763
	0.52 - 0.76	0.8438	0.8438	0.8437	0.8407	0.8384
	0.76 - 1.0	0.8311	0.8312	0.8306	0.8269	0.8236
-	0.28 - 0.52	0.8955	0.8955	0.8929	0.8812	0.8723
	0.52 - 0.76	0.8513	0.8512	0.8497	0.8413	0.8330
	0.76 - 1.0	0.8379	0.8379	0.8359	0.8267	0.8168

with densely packed tracks, the noise reduction afforded by higher thresholds could be beneficial. By filtering out low-energy depositions, mis-identifications of overlapping signals from adjacent tracks might be mitigated. Yet, this advantage ought to be balanced against the observed efficiency loss at higher thresholds. Especially in multi-track events or ones

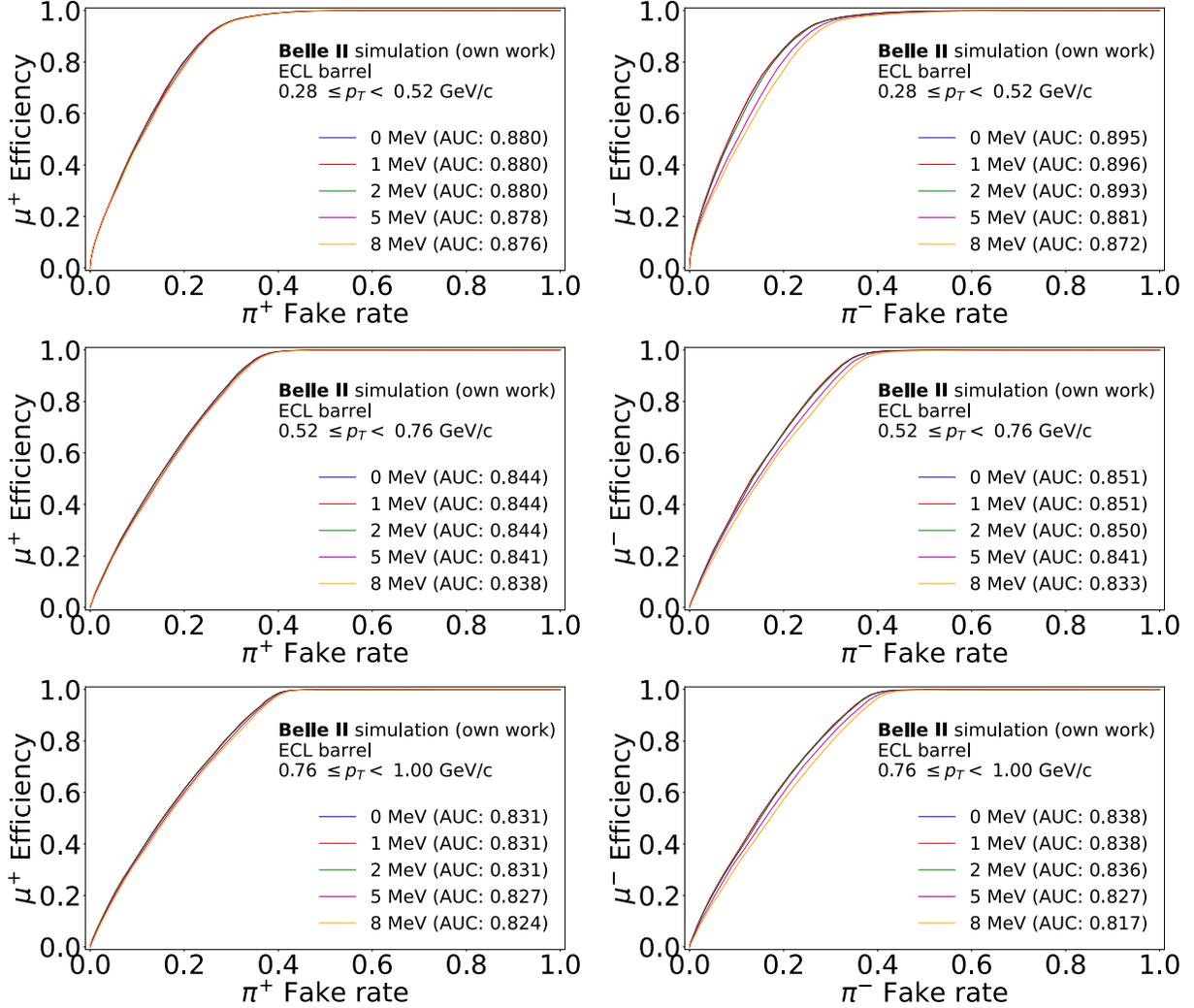


Figure 5.27: ROC curves for positive (left) and negative (right) charged tracks for different energy thresholds across different p_T ranges. From top to bottom: low ($0.28 \leq p_T < 0.52$ GeV/c), intermediate ($0.52 \leq p_T < 0.76$ GeV/c), and high ($0.76 \leq p_T < 1.0$ GeV/c) transverse momentum.

with close-by tracks, a lower threshold could aid in discerning individual particle signatures amidst densely packed signals.

In conclusion, the interplay between energy threshold and model proficiency is not linear. The energy threshold functions as both a filter and a lens, where higher thresholds, though mitigating potential noise, may also discard valuable insights. Performance degradation at higher thresholds might reflect the model's reliance on nuanced signals, which become filtered out with more restrictive thresholds. Even minor threshold alterations engender perceptible performance shifts, especially in challenging phase space regions. Thus, energy threshold deployment in preprocessing warrants prudent consideration to aim for optimal model performance. The notion of adaptive thresholds, which fluctuate based on the local energy landscape of the track or event, might offer a collaborative avenue for

ML practitioners and particle physicists, ensuring that vital signals aren't inadvertently forsaken.

5.8 Beam Backgrounds' Effect

Beam-induced backgrounds, stemming from interactions between primary beams and various machine components, significantly influence the workings of particle detectors. The subsequent noise and potential instrumental inefficiencies can sometimes lead to misinterpretations of experimental results. Given the precision demanded in particle detection, it is paramount to understand and counteract these background effects.

In the ECL, where accuracy is crucial, beam backgrounds can induce additional energy deposits, potentially skewing the measured energy of primary particles. Similarly, in the tracker, the increased rate of false hits from scattered particles due to beam backgrounds can compromise momentum resolution and PID. Furthermore, these backgrounds can also introduce systematic errors into experimental data, which, if unaccounted for, can yield biased results. Thus, understanding and accounting for these backgrounds are essential for ensuring the reliability of experimental findings.

To provide a visual illustration of this phenomenon, 7×7 pixel images of muons and charged pions are analyzed. The emphasis is on observing the effect of varying beam background levels on these images. For this study, muons and charged pions that hit the same crystal position with similar p_T are selected from samples with different beam background levels: $BG \times 0$, $BG \times 1$, $BG \times 2$, and $BG \times 5$. These hits are situated in the heart of the ECL barrel at $\theta_{ID} = 40$ and $\phi_{ID} = 70$, with their p_T s ranging between 0.40 GeV/ c and 0.44 GeV/ c .

Figures [5.28](#) and [5.29](#) show the 7×7 pixel images of μ^+ and π^+ , respectively, under these varying beam backgrounds. As the beam background ascends from $BG \times 0$ (no background) to $BG \times 5$, an evident increase in noise levels can be discerned, reinforcing the challenges posed by higher backgrounds.

5.8.1 Testing on Different Beam Backgrounds

The ROC curves, as presented in Figure [5.30](#), offer a comprehensive view of the classification performance across positive and negative charged particles under varying beam background levels. These curves not only bring to light the inherent difficulties posed by rising beam background noise but also shed light on how the models, trained under specific conditions, fare when faced with these challenges.

For both positive and negative charges, a noticeable trend is the decline in AUC as the beam background intensifies. This is a clear indicator of the diminishing discriminatory power of the models in higher noise environments. Particularly between $BG \times 2$ and $BG \times 5$, the drop is pronounced, suggesting a threshold beyond which the background noise starts having a more adverse impact on classification performance.

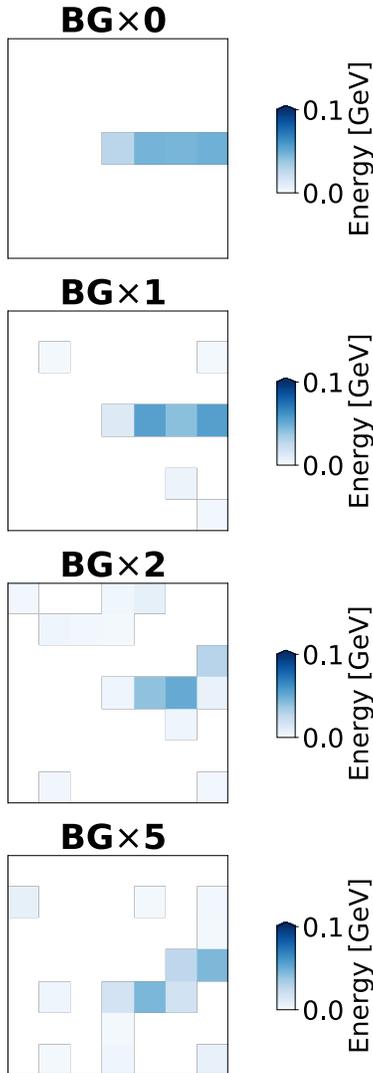


Figure 5.28: 7×7 pixel images of positive muons (μ^+) at various beam background levels: $BG \times 0$ (no beam background), $BG \times 1$, $BG \times 2$, and $BG \times 5$, displayed from top to bottom. An increase in background level corresponds to an increase in observed noise.

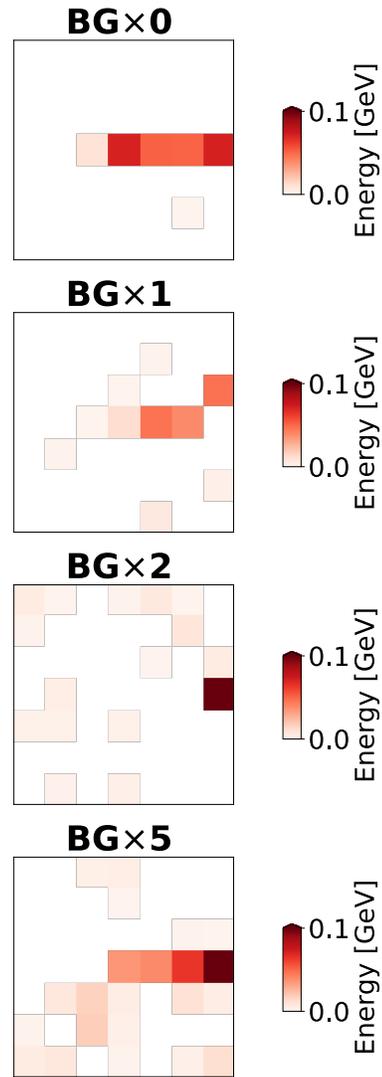


Figure 5.29: 7×7 pixel images of positive pions (π^+) at different beam background levels: $BG \times 0$ (no beam background), $BG \times 1$, $BG \times 2$, and $BG \times 5$, presented from top to bottom. A noticeable increase in noise is observed with the elevation in background levels.

Analyzing the behavior for positive charged tracks, there's a nuanced interplay. The efficiency first sees a rise when moving from $BG \times 0$ to $BG \times 1$, possibly due to the model's ability to leverage certain noise patterns beneficially. However, as the background continues to increase, the efficiency drops, bottoming out at $BG \times 5$. This could be attributed to the overwhelming noise that masks the distinguishing features of positive charged particles.

In contrast, the efficiency curve for negative charged tracks paints a different story. The

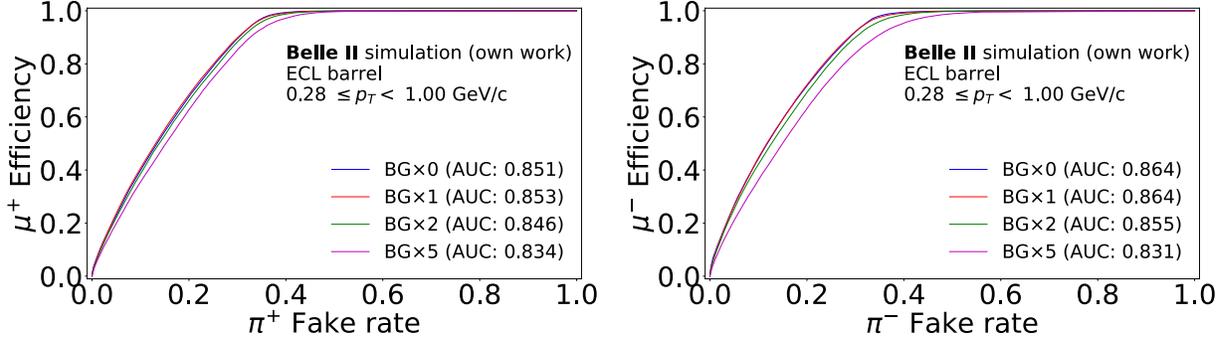


Figure 5.30: ROC curves showing the discriminative performance of the CNN-based PID method across various levels of beam backgrounds trained with $BG \times 1$.

consistent decline with increasing background hints at a more linear relationship between noise and classification performance for these tracks. The sharp drop between $BG \times 2$ and $BG \times 5$ might be indicative of specific interference patterns that particularly obscure the characteristics of negative charges.

The analytical exploration of the models' performance is further enriched by a focused evaluation based on specific operational benchmarks. Table 5.13 and 5.14 encapsulates the essential statistics, detailing muon efficiencies at fixed pion fake rates of 1%, 5%, 10%, and 20%, as well as pion fake rates at fixed muon efficiencies of 80%, 90%, 95%, and 99%. This tabular representation elucidates the nuanced performance variations of the models across different beam background scenarios, providing a granular perspective that complements the overarching insights garnered from the ROC curves. The meticulous breakdown facilitates a more informed and practical understanding of the models' robustness and adaptability, essential for optimizing their application.

5.8.2 Training on Different Beam Backgrounds

In the preceding subsection, the behavior of various $BG \times N$ samples was detailed, focusing on instances where the CNN was trained on $BG \times 1$ and subsequently tested on $BG \times 0$, $BG \times 1$, $BG \times 2$, and $BG \times 5$. Expanding on this investigation, the following delves into the outcomes when the CNN model, maintaining an identical architecture, is trained on samples with varying beam background levels.

Several insights emerge from an examination of the ROC curves in Figures 5.30, 5.31, 5.32, and 5.33 corresponding to different combinations of training and testing on various beam background levels:

- **Consistency Across Charges:** A consistency in the behavior of ROC curves for both positive and negative charged tracks is observed. Such uniformity, while anticipated, serves to confirm the model's consistent response to the inherent properties of tracks, irrespective of their charge. However, discrepancies in absolute ROC values between the two charges are noted, as visualized in Figures 5.34 and 5.35 for positive and negative charged tracks, respectively.

Table 5.13: μ^+ efficiencies at fixed π^+ fake rates (top) and π^+ fake rates at fixed μ^+ efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/c for various beam background levels. Fix values 1 %, 5 %, 10 %, and 20 % are selected for π^+ fake rates and 80 %, 90 %, 95 %, and 99 % for μ^+ efficiencies.

Beam background level	μ^+ efficiency at fixed π^+ fake rates (%)			
	1 %	5 %	10 %	20 %
BG \times 0	7.21	25.71	44.48	74.45
BG \times 1	6.80	23.57	40.63	68.41
BG \times 2	6.37	22.16	38.62	66.24
BG \times 5	5.09	20.99	37.49	65.08

Beam background level	π^+ fake rate at fixed μ^+ efficiencies (%)			
	80 %	90 %	95 %	99 %
BG \times 0	22.13	26.57	29.60	34.82
BG \times 1	25.09	30.11	33.32	39.34
BG \times 2	26.08	31.00	34.12	40.38
BG \times 5	26.38	31.41	34.62	42.75

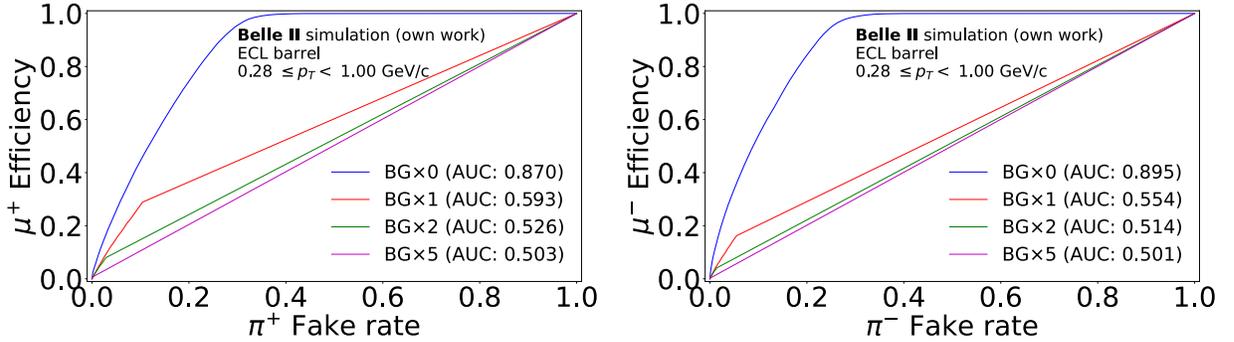


Figure 5.31: ROC curves showing the discriminative performance of the CNN-based PID method across various levels of beam backgrounds trained with BG \times 0.

- Model Robustness and Adaptability:** It is observed that the model’s performance reaches its zenith when tested on the same beam background on which it is trained. This phenomenon underscores the model’s ability to adapt to specific noise and signal patterns within that particular background. Yet, with the amplification of the beam background during testing, a decline in the model’s performance becomes evident. This is particularly pronounced when there exists a disparity between the training and testing backgrounds. Such a decline can be attributed to the increased

Table 5.14: μ^- efficiencies at fixed π^- fake rates (top) and π^- fake rates at fixed μ^- efficiencies (bottom) in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/c for various beam background levels. Fix values 1 %, 5 %, 10 %, and 20 % are selected for π^- fake rates and 80 %, 90 %, 95 %, and 99 % for μ^- efficiencies.

Beam background level	μ^- efficiency at fixed π^- fake rates (%)			
	1 %	5 %	10 %	20 %
BG \times 0	11.01	33.78	54.01	84.29
BG \times 1	8.18	26.59	44.64	72.07
BG \times 2	7.22	24.74	41.92	69.72
BG \times 5	5.90	21.63	38.26	65.69

Beam background level	π^- fake rate at fixed μ^- efficiencies (%)			
	80 %	90 %	95 %	99 %
BG \times 0	18.34	22.39	25.12	30.65
BG \times 1	23.46	28.68	32.07	39.09
BG \times 2	24.46	29.80	33.25	41.01
BG \times 5	26.20	31.43	35.24	45.10

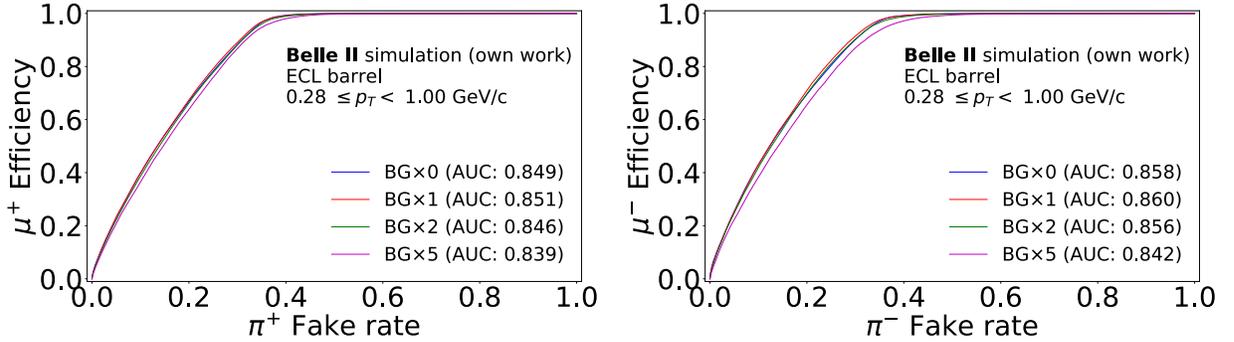


Figure 5.32: ROC curves showing the discriminative performance of the CNN-based PID method across various levels of beam backgrounds trained with BG \times 2.

background noise potentially overshadowing the signals, making the discrimination task more arduous, as illustrated in the corresponding ROC plots.

- **Interplay of Physics and Geometry:** The disparity in performance across different beam backgrounds accentuates the crucial role of the ECL's geometry in tandem with intrinsic physics processes. With an augmented beam background, a surge in the number of events within the ECL is recorded, resulting in intricate spatial patterns of energy deposits. Such complex patterns of energy dispersion might impede

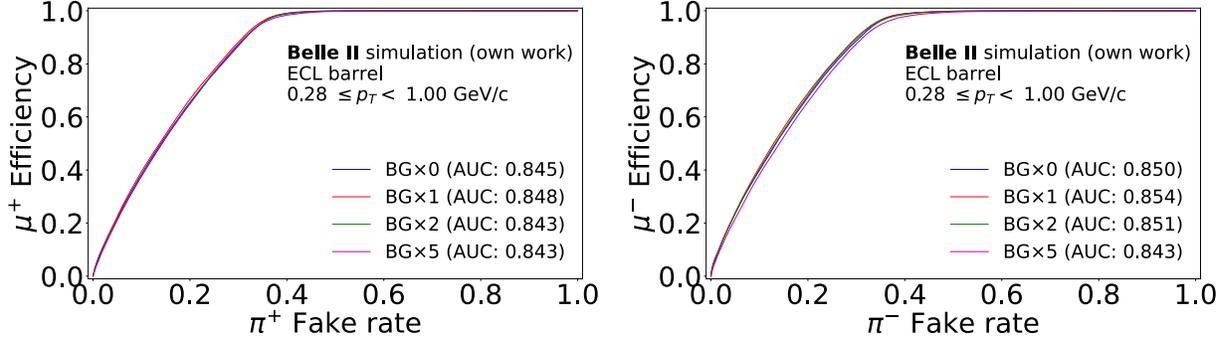


Figure 5.33: ROC curves showing the discriminative performance of the CNN-based PID method across various levels of beam backgrounds trained with $BG \times 5$.

the model’s ability to differentiate between pion and muon tracks, especially when not trained on those specific patterns.

- **Diverse Background Training:** A noteworthy resilience is exhibited by the model when trained on beam backgrounds with heightened noise levels, such as $BG \times 5$. When subjected to tests on varied beam backgrounds, this model demonstrates a robust adaptability, potentially indicating its exposure to a plethora of intricate patterns during training. However, it is emphasized that such adaptability might not always translate to optimal performance for specific isolated backgrounds.

In conclusion, while training a model on a specific background can optimize performance for that particular background, it may not guarantee optimal performance across varying backgrounds. Therefore, understanding the interactions between various beam backgrounds and the CNN model’s training becomes imperative for achieving optimal performance.

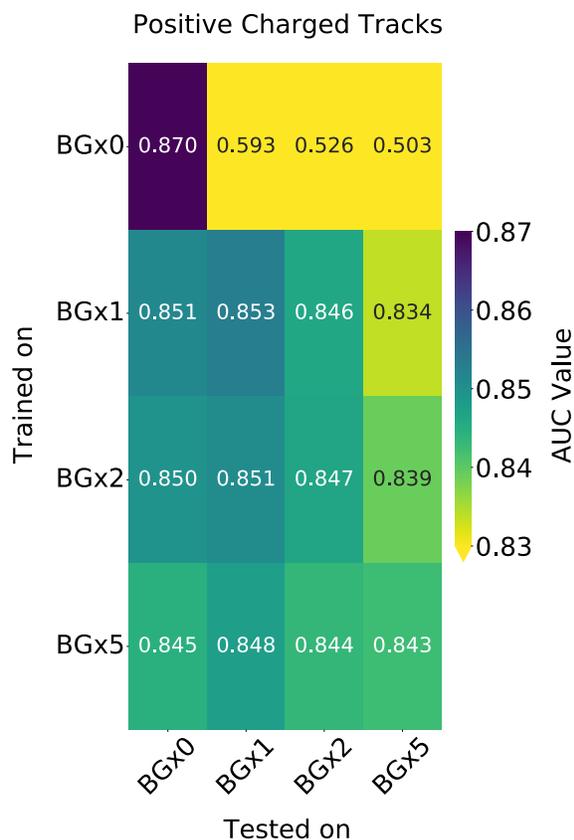


Figure 5.34: Heatmap illustrating the AUC values for the CNN model trained and tested on various beam background levels for positive charged tracks. The model is trained on beam background levels denoted by row labels and tested on the levels specified by column headers. The color intensity represents the AUC value, with deeper colors indicating higher performance. It is evident from the heatmap that the model’s performance is optimal when trained and tested on the same beam background level, with performance generally declining as the discrepancy between training and testing backgrounds increases.

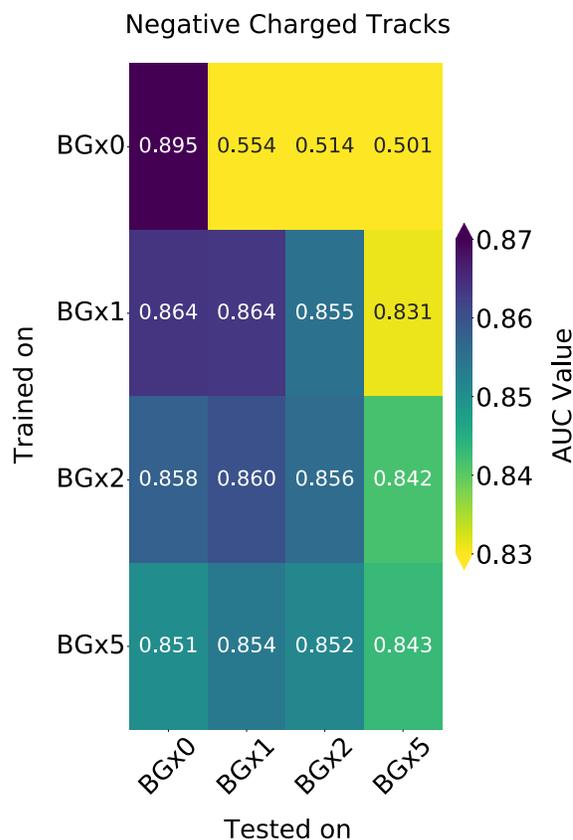


Figure 5.35: Heatmap illustrating the AUC values for the CNN model trained and tested on various beam background levels for negative charged tracks. The model is trained on beam background levels denoted by row labels and tested on the levels specified by column headers. The color intensity represents the AUC value, with deeper colors indicating higher performance. It is evident from the heatmap that the model’s performance is optimal when trained and tested on the same beam background level, with performance generally declining as the discrepancy between training and testing backgrounds increases.

Chapter 6

CNN Validation with Data

In particle physics, controlled simulated data often serves as the foundational platform for model development. However, the true test of a model’s mettle is its performance on data, which embodies the multifaceted intricacies of physical processes. This chapter evaluates the CNN model against two distinct datasets: collision dataset (data) and MC simulation. The latter meticulously mirrors real experimental conditions and thereby offering a rigorous assessment environment for the model.

The focus of this evaluation is twofold. First, muons are selected from radiative dimuon events, $e^+e^- \rightarrow \mu^+\mu^-\gamma$, providing a clean muon sample. In parallel, charged pions are selected in decays of D^* mesons, $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$. By analyzing the model’s performance on these specific channels, its precision and adaptability to the actual experimental environment is assessed.

The overarching goal of this chapter is not just to measure the CNN’s efficacy on these datasets, but also to garner insights into its strengths and potential areas for PID improvements. Such insights are pivotal for future model iterations and its broader applicability.

6.1 Collision Dataset

This study uses a dataset of e^+e^- collisions produced by the SuperKEKB accelerator and collected by the Belle II detector in 2020 and 2021. The on-resonance collision dataset corresponds to an integrated luminosity of $\int L dt = 21.5 \text{ fb}^{-1}$. On-resonance data are collected at a collision energy of $\sqrt{s} \approx 10.58 \text{ GeV}$, corresponding to the $\Upsilon(4S)$ resonance threshold.

6.2 MC Simulated Samples

MC simulations are indispensable tools in the realm of experimental particle physics, working hand-in-hand with collision data. They are essential in analyses for providing a theoretical framework and reference against which experimental data can be compared and

interpreted. These simulations provide a controlled setting to test theoretical models and predict detector responses. Crucial for calibrating and correcting detectors, they also estimate and discern background noise that may mask key signals. Data collected is routinely compared with MC simulation predictions. Discrepancies highlight areas for theoretical refinement or detector model improvement. Essentially, MC simulations are key to interpreting and analyzing experimental data with greater accuracy and efficiency.

Table 6.1 details the simulated signal samples. While these simulations are based on the anticipated operating conditions of SuperKEKB during data collection, they do not factor in the variations in actual background rates across different run periods. Following the simulation phase, machine background measurements are overlaid onto each MC sample.

Table 6.1: MC signal samples used for $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$ and $e^+e^- \rightarrow \mu^+\mu^-\gamma$ channels.

Sample Name	Generator	$\int L dt$ [fb ⁻¹]	Channels
continuum $c\bar{c}$	KKMC [87] + PYTHIA 8 [88] + EvtGen [89]	10	$D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$
$\mu\mu(\gamma)$	KKMC [87]	100	$e^+e^- \rightarrow \mu^+\mu^-\gamma$

6.3 $e^+e^- \rightarrow \mu^+\mu^-\gamma$

Events involving radiative dimuons serve as the benchmark for assessing the efficiency of muon identification, offering the advantage of isolating highly pure samples of muons.

6.3.1 Preselection Criteria

For the $e^+e^- \rightarrow \mu^+\mu^-\gamma$ channel, a distinct preselection grounded in the high-level, software-based trigger (HLT) reconstruction is applied, optimized for selecting radiative dimuon events. The criteria mandate that events can encompass a maximum of three track candidates fulfilling the following conditions:

- $|dr| < 2$ cm,
- $|dz| < 4$ cm,
- $p_T > 0.2$ GeV/ c .

Here, dr and dz represent the proximities of the track's closest approach to the IP within the $r - \phi$ plane and along the z -axis, respectively. Each of these tracks should

register at least a single hit in the CDC or KLM and correlate with an ECL cluster energy $E_{\text{cluster}} < 0.4$ GeV. In these events, there must be a pair of tracks consistent with the kinematic properties of a dimuon pair, meaning they must have a recoil system momentum $p_{\text{recoil}} > 0.1$ GeV/ c , compatible with the emission of a single photon, and $E_{\text{cluster}} < 0.25$ GeV for each track, with $|\Delta\phi_{\text{cluster}} \geq \pi/2|$. Lastly, the track with the highest momentum in this pair must satisfy $p_{\text{lab}} > 1$ GeV/ c and the other $p_{\text{lab}} < 3$ GeV/ c .

6.3.2 Event Selection

From the skimmed samples, candidates are selected based on specific criteria. The selection requires exactly two tracks with opposing charges with conditions $|dr| < 2$ cm and $|dz| < 5$ cm. Photon candidates must have a minimum energy of 1.5 GeV ensuring it is captured within the ECL acceptance ($-0.8660 < \cos(\theta_{\text{cluster}}) < 0.9563$). The analysis is further refined by imposing a constraint on the total invariant mass of the $\mu^+\mu^-\gamma$ system.

6.3.3 Data and MC Comparison

In order to ensure that the selection are consistent between data and MC, a comparison is made between data and MC signal.

A tag and probe method is used for the estimation of muon efficiency. The presence of a so-called tag muon corresponding to a track satisfying a tight PID requirement of $\text{muonID} > 0.9$, then the other muon is considered as the probe candidate [90]. The muonID variable represents a likelihood ratio which include contributions from all sub-detectors.

The backgrounds for this channel include $\pi^+\pi^-\gamma$, $K^+K^-\gamma$, $\tau^+\tau^-$, e^+e^- , and $e^+e^-\mu^+\mu^-$. The total invariant mass includes large contribution from $\tau^+\tau^-$ in the low mass region of the invariant mass spectrum. The contribution from $\tau^+\tau^-$ is removed with a proper selection on the total invariant mass: $9 < M_{\mu^+\mu^-\gamma} < 10.8$ GeV/ c^2 . With this selection most of the backgrounds can be removed. However there are still contributions from Bhabha events ($e^+e^- \rightarrow e^+e^-$) which is removed via tagging one muon with $\text{muonID} > 0.9$. After applying the tag and probe selection, no background remains from $e^+e^-\mu^+\mu^-$.

The distribution of total invariant mass, $M_{\mu^+\mu^-\gamma}$, for signal sample after aforementioned selections and tagging μ^+ is shown in Figure 6.1. The MC distribution has been normalised to the collision dataset integrated luminosity.

Because of an observed bias in the magnetic field map used during data processing, a global scaling factor of 0.9997, derived from the calibration of the invariant mass peak of the D^0 meson, is applied to correct the muon momentum in the data.

The distribution of $M_{\mu^+\mu^-\gamma}$ indicates an average difference of approximately 6% between the data and the MC. This difference is most pronounced in the KLM endcap regions and for tracks with very high momentum ($p_{\text{lab}} > 6$ GeV/ c). The cause of this discrepancy is attributed to the lack of momentum scale corrections as a function of the track's polar angle and are specifically fine-tuned for very high momentum tracks.

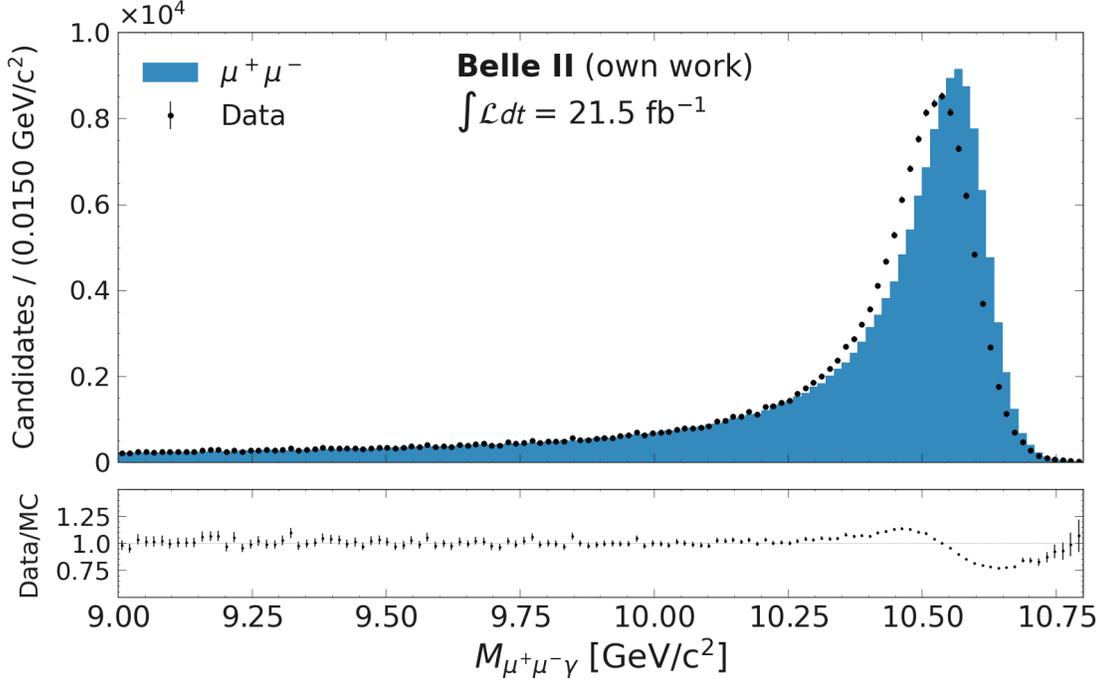


Figure 6.1: Distribution of the $M_{\mu^+\mu^-\gamma}$ invariant mass for $e^+e^- \rightarrow \mu^+\mu^-\gamma$ candidates when μ^+ is selected as tag ($\text{muonID} > 0.9$) and μ^- as probe track. The MC signal (represented as $\mu^+\mu^-$) is normalised to the integrated luminosity of 21.5 fb^{-1} .

The focus of this study is the validation of the CNN models on the extrapolated tracks passing through the ECL barrel with transverse momentum between 0.28 and 1.0 GeV/c. To have a better understanding about CNN input distributions, the data and MC distributions of p_T , θ_{ID} , and ϕ_{ID} of μ^+ and μ^- are shown in Figure [6.2](#), [6.3](#), and [6.4](#), respectively. The MC distribution has been normalised to the collision dataset integrated luminosity.

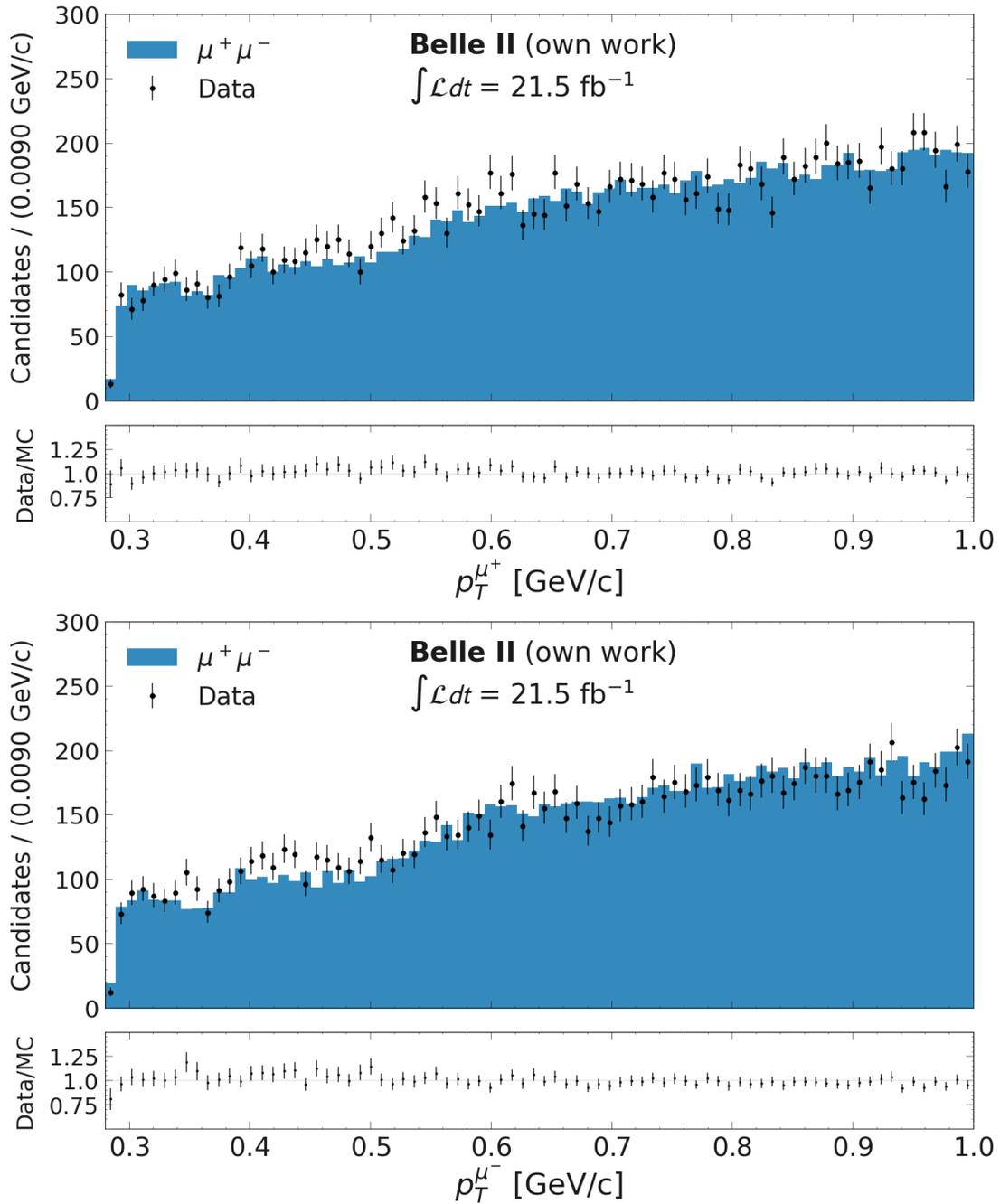


Figure 6.2: Distribution of p_T of μ^+ (top) and μ^- (bottom) after applying $\text{muonID} > 0.9$ on the tag track. The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

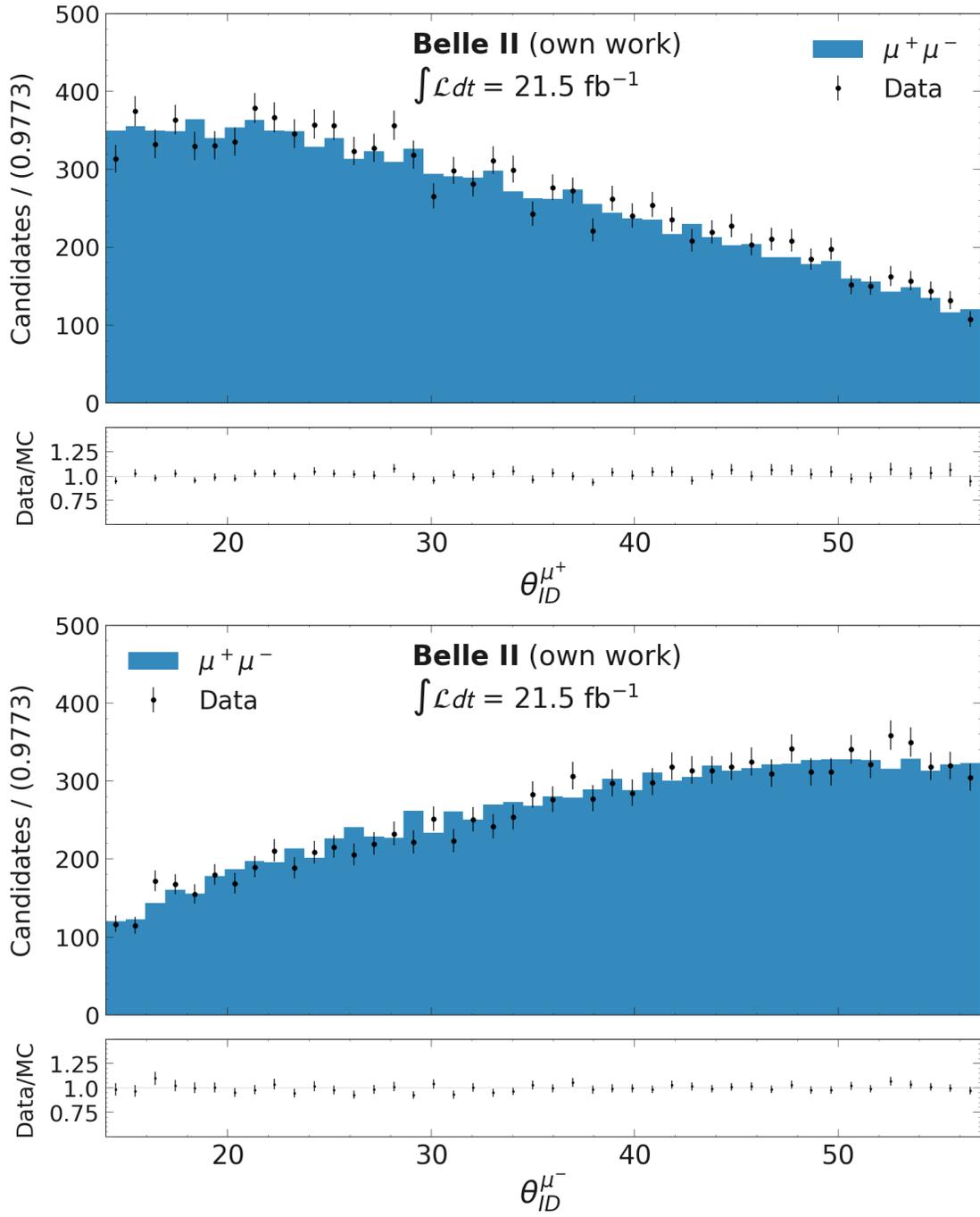


Figure 6.3: Distribution of θ_{ID} of μ^+ (top) and μ^- (bottom) after applying $\text{muonID} > 0.9$ on the tag track. The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

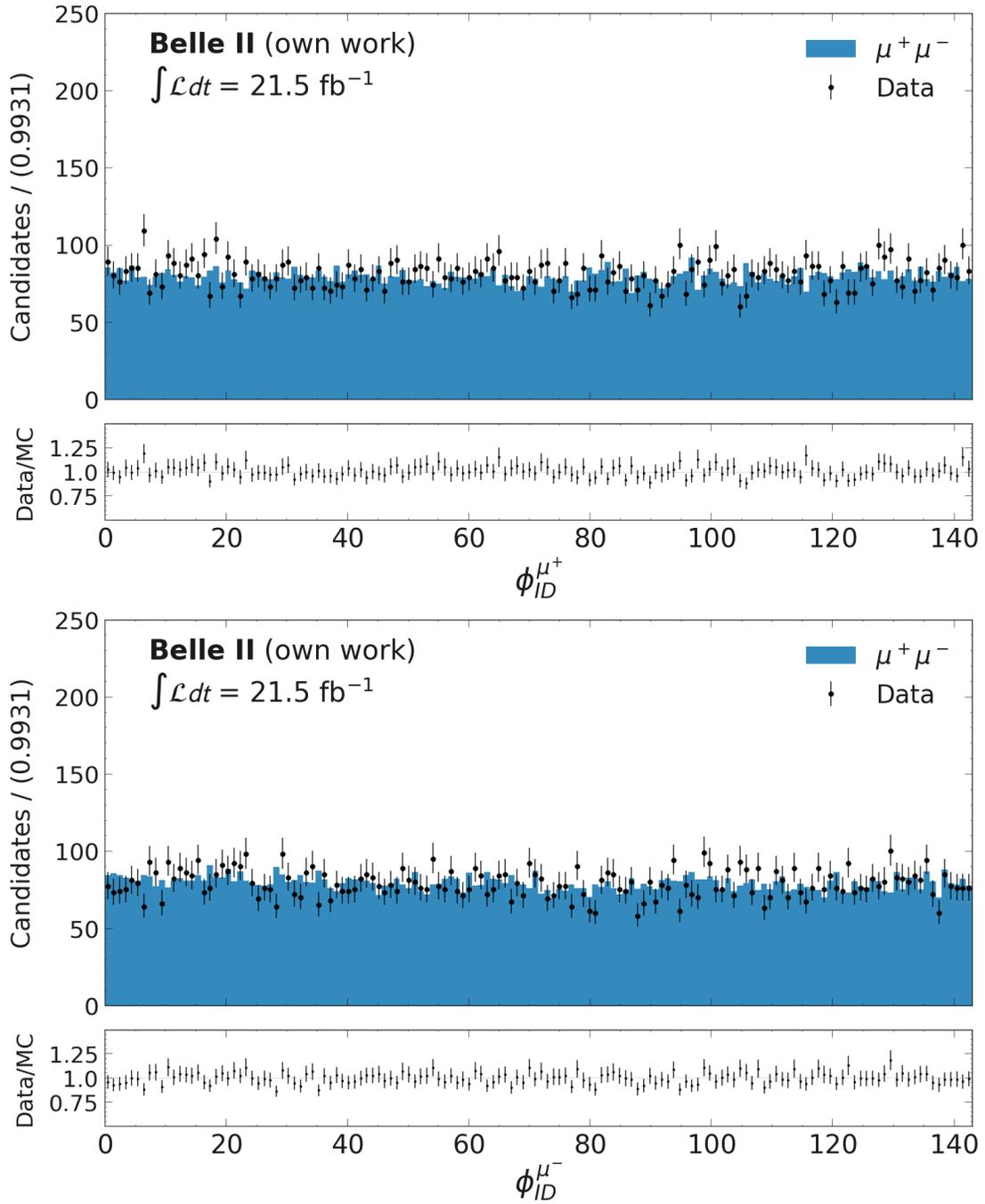


Figure 6.4: Distribution of ϕ_{ID} of μ^+ (top) and μ^- (bottom) after applying $\text{muonID} > 0.9$ on the tag track. The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

6.4 $D^{*+} \rightarrow D^0(\rightarrow K^- \pi^+) \pi^+$

The $D^{*+} \rightarrow D^0(\rightarrow K^- \pi^+) \pi^+$ candidates are used to measure the the probability to mis-identify a pion as a lepton. The D^{*+} meson has a mass of $2.01 \text{ GeV}/c^2$ and it decays into $D^0 \pi^+$ with a branching fraction of 67.7% [91]. The D^0 meson, with a mass of $1.86 \text{ GeV}/c^2$, consequently decays into $K^- \pi^+$ with a branching fraction of 3.89% [91]. The decay chains of D^{*+} and D^{*-} are:

$$D^{*+} \rightarrow D^0[\rightarrow K^- \pi^+] \pi^+, \quad (6.1)$$

$$D^{*-} \rightarrow \bar{D}^0[\rightarrow K^+ \pi^-] \pi^-. \quad (6.2)$$

6.4.1 Preselection Criteria

For the D^{*+} decay channel, a preselection is employed rooted in the HLT reconstruction. This is specifically tailored to preserve events stemming from hadronic B decays. Such events must possess at least three tracks satisfying the following conditions:

- $|dr| < 2 \text{ cm}$,
- $|dz| < 4 \text{ cm}$,
- $p_T > 0.2 \text{ GeV}/c$.

6.4.2 Event Selection

From the skimmed samples, candidates are selected based on specific criteria. To ensure tracks emerge near the IP, criteria $|dr| < 2 \text{ cm}$ and $|dz| < 4 \text{ cm}$ are enforced. D^0 mesons are identified from pairs of tracks with opposing charges, and the presence of a third, momentum-restricted track, termed the *slow pion*, serves to tag these D^0 mesons. To specifically isolate candidates originating from the $e^+e^- \rightarrow c\bar{c}$ continuum process, the D^+ momentum in the CM frame should exceed $p_{D^{*+}} > 2.5 \text{ GeV}/c$. Furthermore, a mass window restriction for the $D^0 - D^{*+}$ difference is set at $|\Delta M - 0.1453| < 0.0015 \text{ GeV}/c^2$. Bypassing the Cabibbo-suppressed D^0 decays, the *slow pion*'s charge is inversely related to the kaon's charge from the D^0 decay. This relationship facilitates the kaon's identification and the recognition of the pion with the opposing charge, eliminating the need for additional selections.

6.4.3 Data and MC Comparison

In the process of validating the CNN models, particularly for muon/pion separation, a comprehensive analysis of the D^{*+} decay channel is conducted. The mass distribution of the D^0 candidates is fitted with a model that consists of two Gaussian functions for the signal and a third-order Chebychev polynomial for the background component. This model

Table 6.2: Post-fit values of the model for the D^0 mass distribution from data and MC. The parameters are obtained using a composite model of two Gaussian distributions for the signal and a third-order Chebychev polynomial for the background. The uncertainties represent one standard deviation from the fitted values.

Parameter	Data	MC
Mean	$1.864722 \pm 0.000017 \text{ GeV}/c^2$	$1.864739 \pm 0.000021 \text{ GeV}/c^2$
$n_{\text{Gauss 1}}$	$(9.12 \pm 0.04) \times 10^4$	$(4.752 \pm 0.028) \times 10^4$
$\sigma_{\text{Gauss 1}}$	$0.004620 \pm 0.000020 \text{ GeV}/c^2$	$0.004226 \pm 0.000023 \text{ GeV}/c^2$
$n_{\text{Gauss 2}}$	$(1.78 \pm 0.05) \times 10^4$	$(7.77 \pm 0.29) \times 10^3$
$\sigma_{\text{Gauss 2}}$	$0.014620 \pm 0.000020 \text{ GeV}/c^2$	$0.014226 \pm 0.000023 \text{ GeV}/c^2$
$n_{\text{Background}}$	$(2.576 \pm 0.031) \times 10^4$	$(7.68 \pm 0.17) \times 10^3$

is applied to datasets obtained from data and MC, which emulate the $c\bar{c}$ signal events. The fit parameters extracted from both data and simulation are presented in Table 6.2.

In the plots in Figure 6.5 illustrating the fits, the mass distributions are shown with the composite model overlaid. The fitted mean positions of the Gaussian components for both data and MC are consistent with the world-average D^0 mass of $1.86 \text{ GeV}/c^2$ [92]. Such consistency is indicative of the mass scale calibration of the detector being accurately reflected in both data and MC.

It is noted that the core signal, represented by the first Gaussian component, exhibits a broader width in the data compared to the simulation. This discrepancy is often expected since simulations may not account for all detector effects that contribute to the width of the mass peak.

The yield of the signal and the non-peaking component, modeled by the Chebychev polynomial, differ between the data and the MC simulation. This is anticipated, as the MC simulation does not include the complete range of background processes present in the actual data. The non-peaking component in the MC is necessitated to account for events that are part of the signal process but do not contribute to the sharp peak.

In the comparative plot as presented in Figure 6.6, a discrepancy is noted in the peak magnitudes, with the MC simulation exhibiting a more pronounced peak than the data. The data/MC ratio plot, situated beneath the main distribution, quantifies these discrepancies. The ratio exceeds unity in the region around the D^0 mass peak, which might suggest an overestimation of the signal in the MC or an under-representation of the background in the data. Such observations are critical as they may point to potential issues in the normalization process or in the modeling of the signal within the MC simulation.

It is imperative to remark that, despite the current focus being the validation of peak positions and overall consistency between data and MC, the noted variance in peak heights

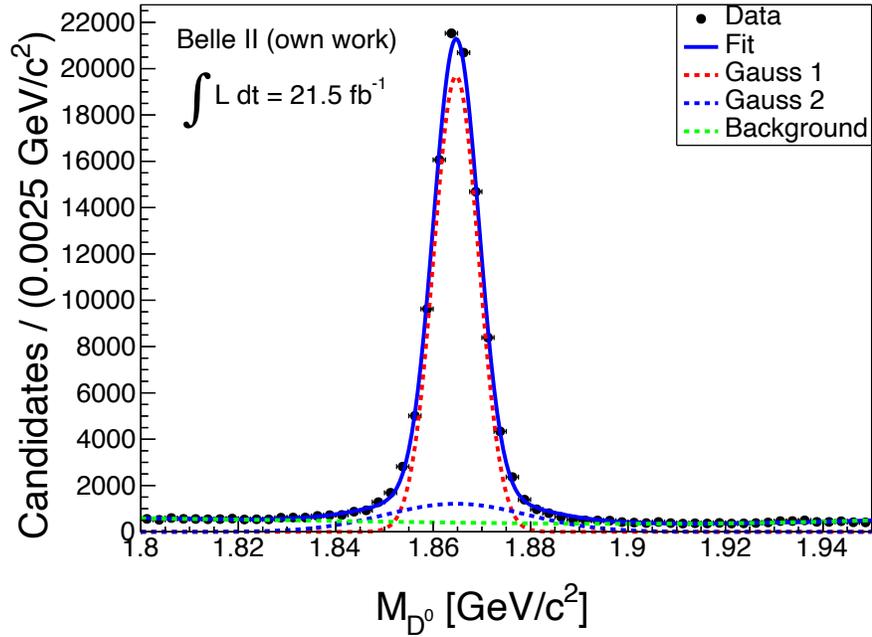
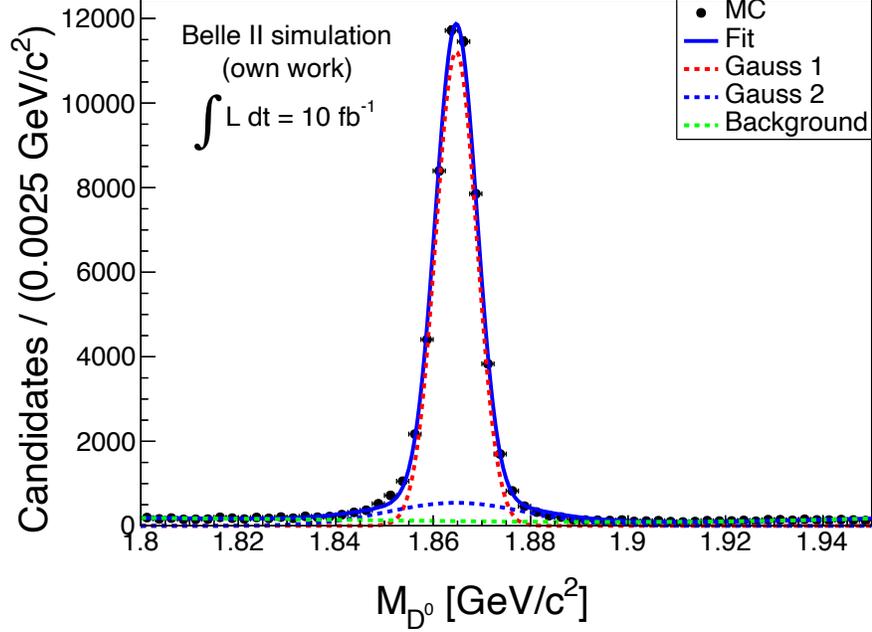


Figure 6.5: Comparison of the fitted D^0 mass distributions for MC (top) and data (bottom). Each plot shows the mass spectrum overlaid with a composite model consisting of two Gaussian components representing the signal and a third-order Chebychev polynomial for the background. These fits validate the consistency of the peak positions between MC and data, crucial for the selection of pure charged pions for subsequent PID studies.

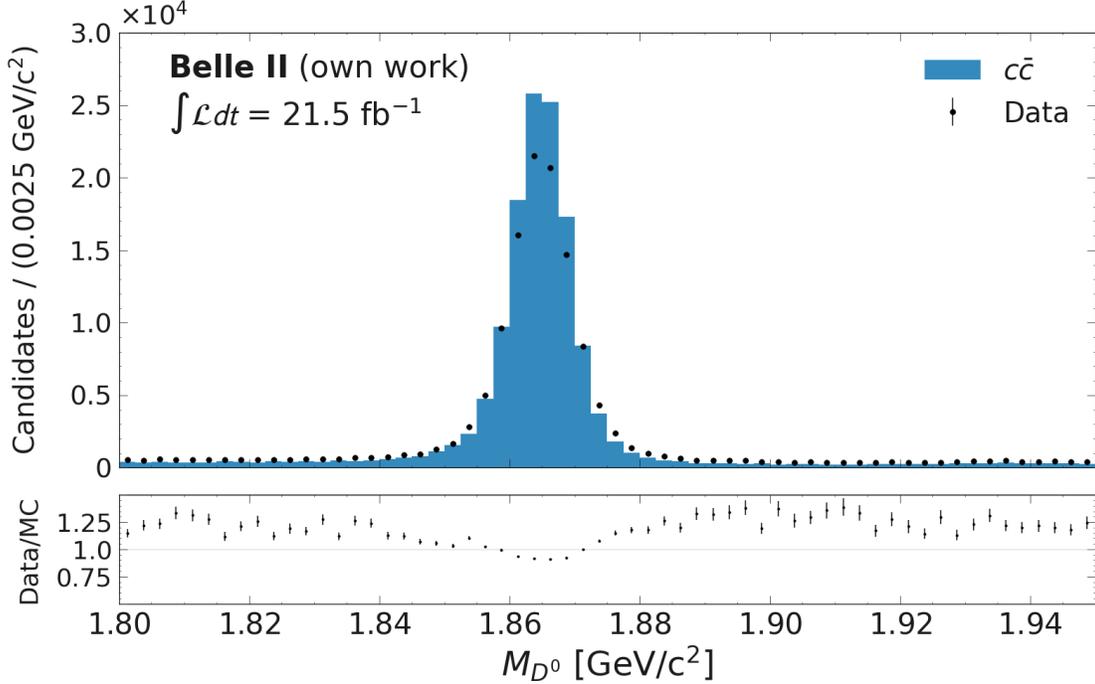


Figure 6.6: Mass distribution of D^0 mesons in the MC simulation of $c\bar{c}$ events, overlaid with data for comparison. The distribution shows the mass peak along with the Data/MC ratio, providing insight into the relative agreement between observed data and theoretical predictions.

could serve as a basis for future detailed studies into systematic uncertainties. Such analyses are beyond the scope of the present work but are acknowledged as vital for subsequent precision measurements and for a comprehensive understanding of the detector’s performance and the underlying physics processes.

As for the validation with $e^+e^- \rightarrow \mu^+\mu^-\gamma$ events, the focus of this study is on validation of the CNN models on the extrapolated tracks passing through the ECL barrel with transverse momentum between 0.28 and 1.0 GeV/c . To have a better understanding about CNN inputs’ distributions, the data and MC distributions of p_T , θ_{ID} , and ϕ_{ID} of π^+ and π^- originating from D^* and D^0 are shown in Figure [6.7](#), [6.8](#), and [6.9](#), respectively. The MC distribution has been normalised to the collision dataset integrated luminosity.

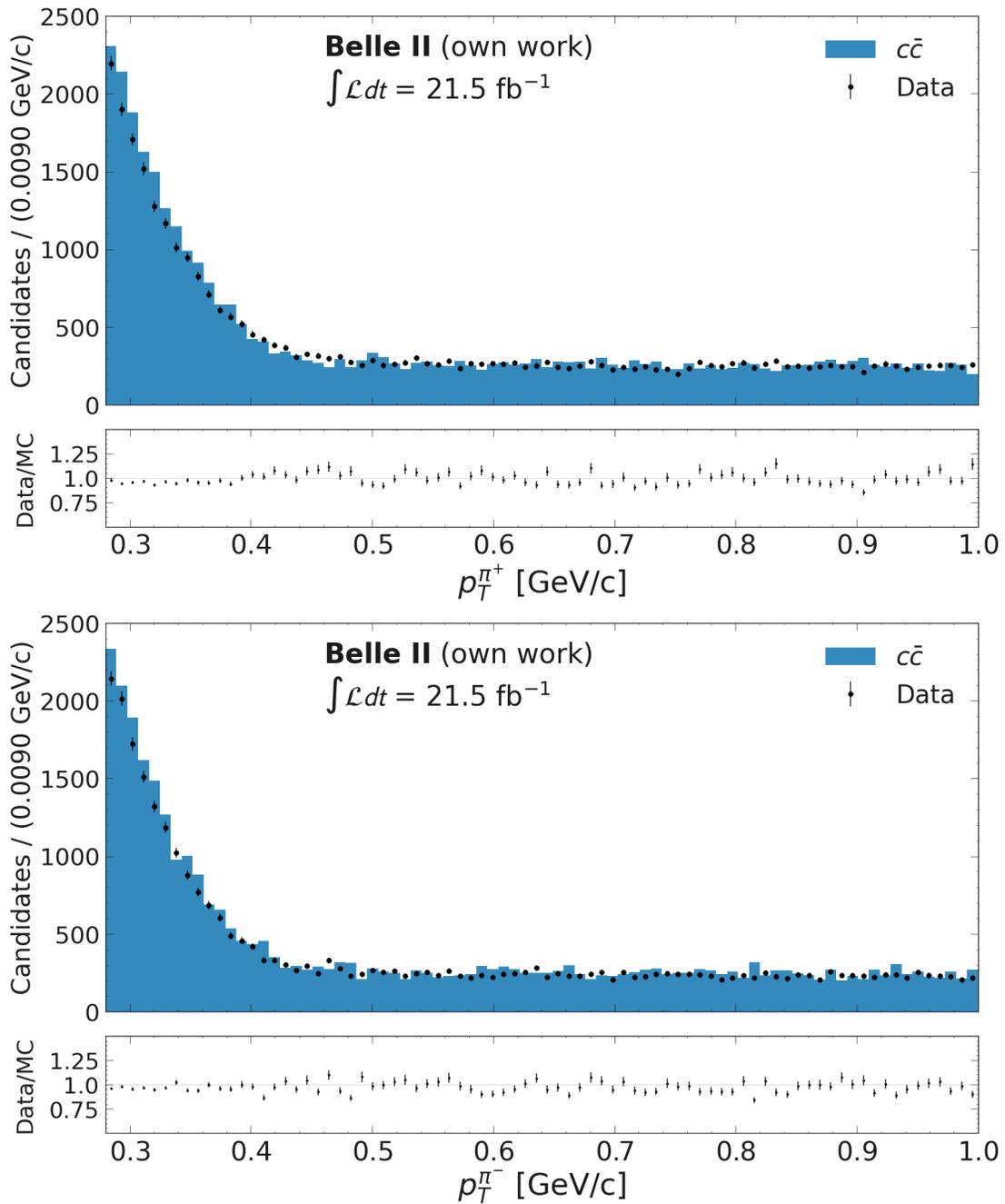


Figure 6.7: Distribution of p_T of π^+ (top) and π^- (bottom). The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

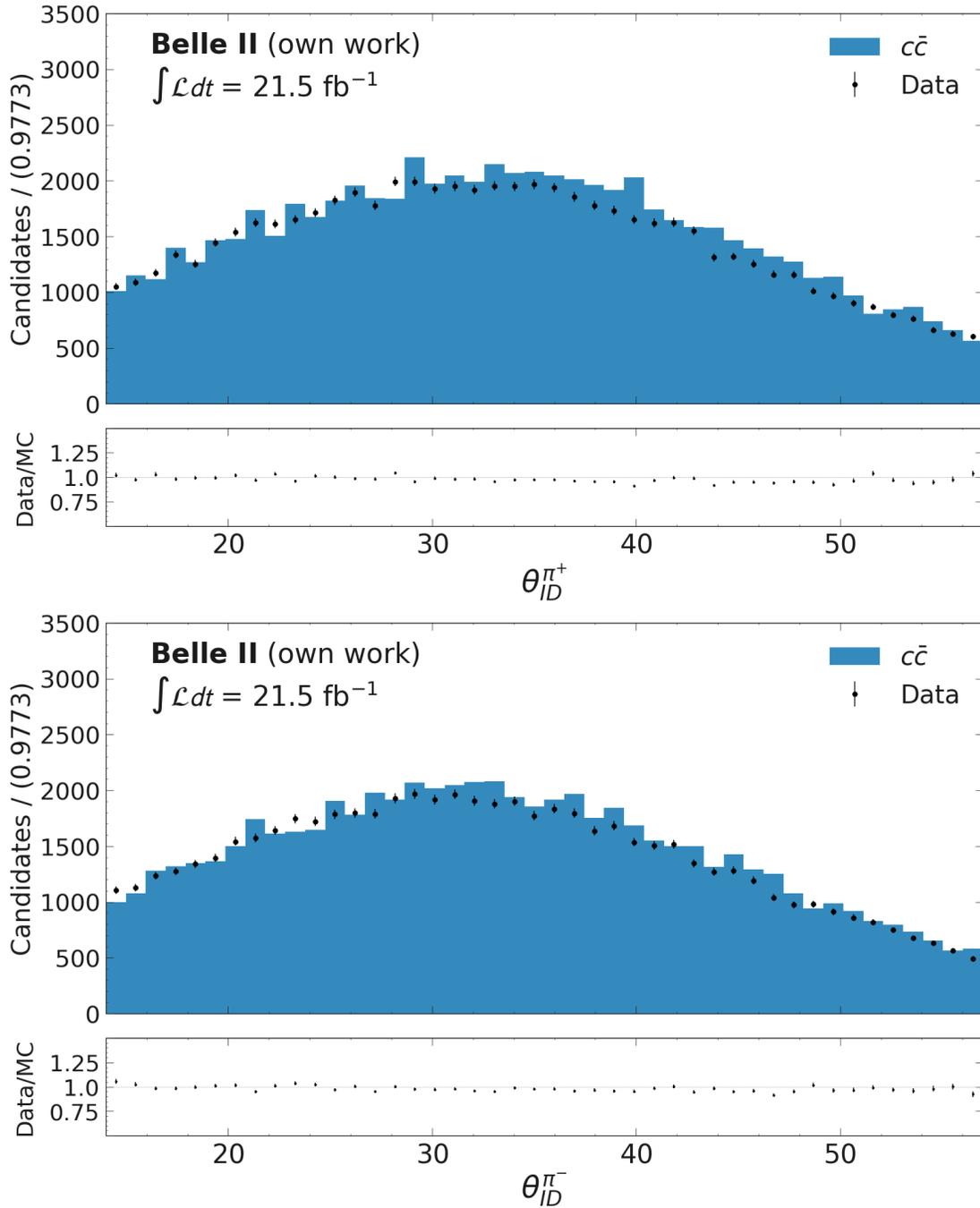


Figure 6.8: Distribution of θ_{ID} of π^+ (top) and π^- (bottom). The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

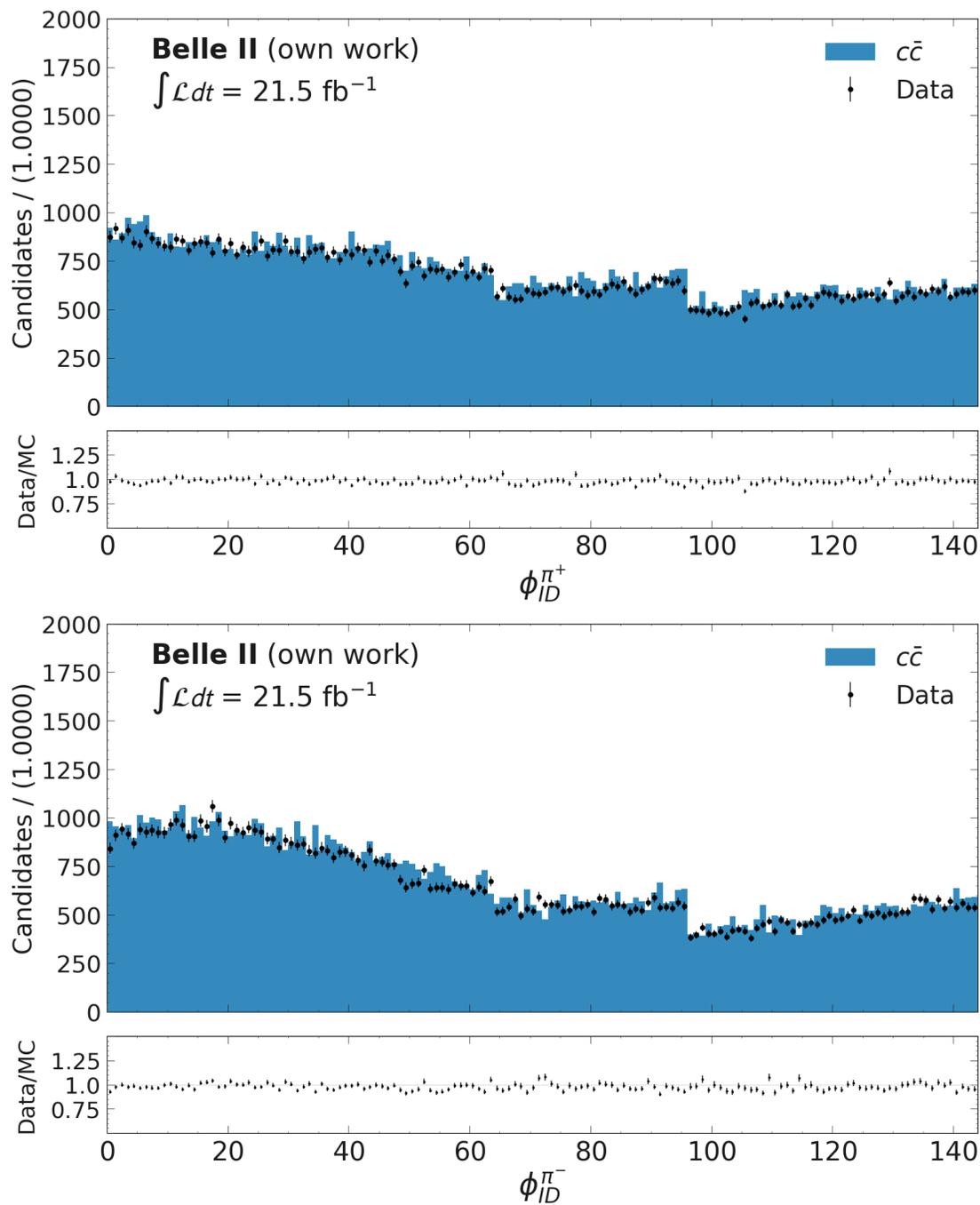


Figure 6.9: Distribution of ϕ_{ID} of π^+ (top) and π^- (bottom). The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

6.5 Comparison of CNN Outputs

The comparison of CNN outputs is crucial for verifying the model’s performance in processing signals from both data and MC. The CNN models for positive and negative charged tracks are trained on samples with nominal beam background ($BG \times 1$) with energy threshold of 1 MeV. The CNN models are exactly the same as used in the previous chapter. However, in this case, the focus is on muons from the radiative dimuon channel and pions from the D^* decay. An overview of the number of muons and pions passing the aforementioned selection described in the previous sections is presented in Table [6.3](#).

Table 6.3: Comparison of μ^\pm and π^\pm counts between data and MC

Particle	Data	MC
μ^+	11 614	53 226
μ^-	11 361	53 283
π^+	25 208	12 125
π^-	24 237	11 685

The outputs of the CNN for μ^+ and μ^- are shown Figure [6.10](#). Both data and MC show a pronounced feature at zero, which is consistent across the various datasets. This feature is attributed to the influence of nearby photons and will be the subject of a detailed examination in subsequent sections. The alignment of this feature in the data and MC histograms indicates that the MC effectively replicates this aspect of the data’s signal characteristics.

Figure [6.11](#) presents CNN output for π^+ and π^- . The plots reveal specific patterns in the CNN’s response to pions, which are consistent with the expected signal profiles. The close correspondence between the data and MC outputs in these channels reinforces the confidence in the MC’s ability to model the pion signals accurately.

The comparison between the CNN outputs for data and MC within each particle type provides insight into the network’s ability to handle data. The degree of agreement between the data and MC outputs suggests that the network has learned features that are not only present in the simulated environment but also manifest in the actual data.

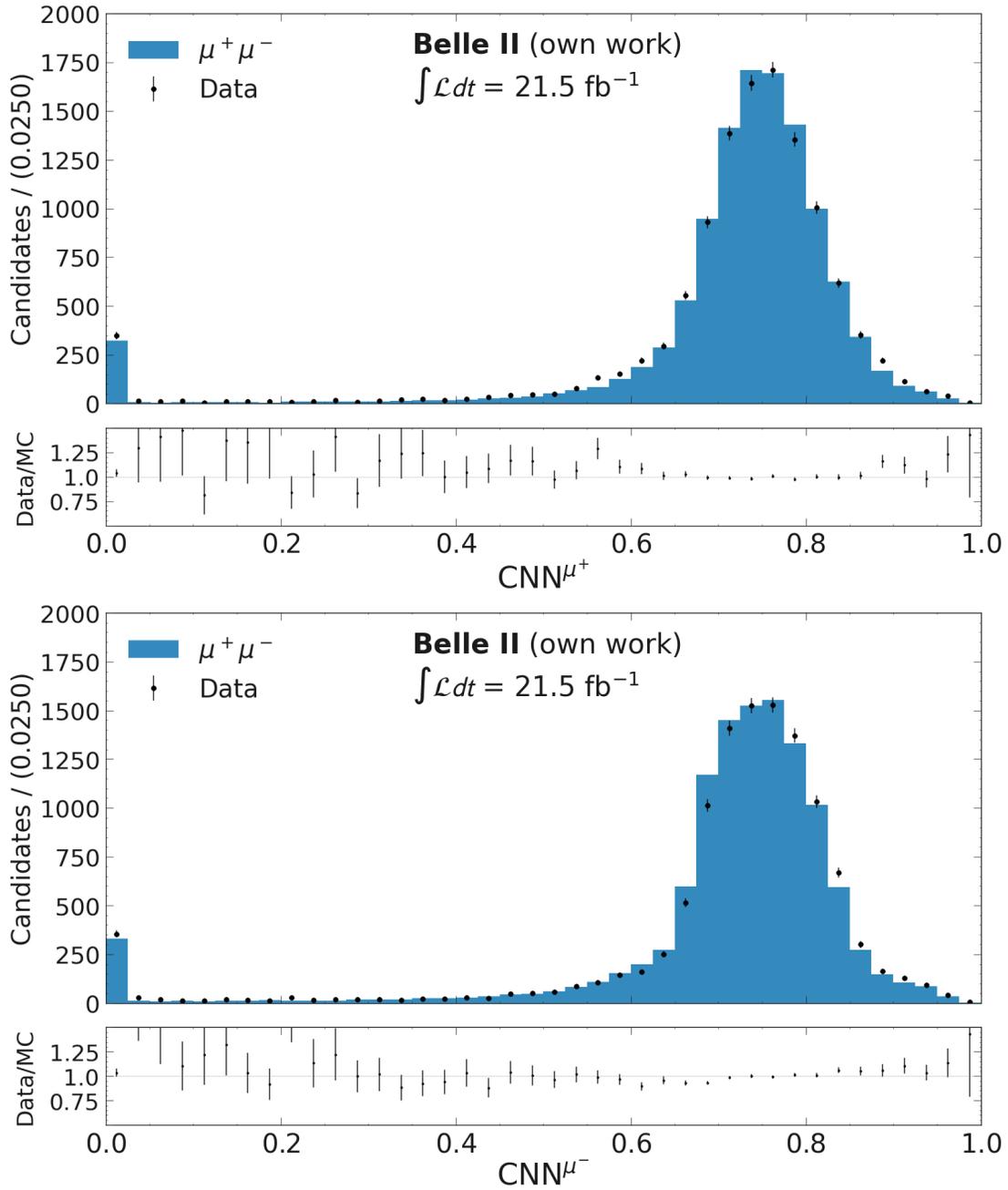


Figure 6.10: Distribution of CNN outputs for μ^+ (top) and μ^- (bottom) from the radiative dimuon channel. The histograms compare the signal response in data to that of the MC simulation. A notable feature at zero is present in both distributions, which is attributed to the influence of nearby photons. The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

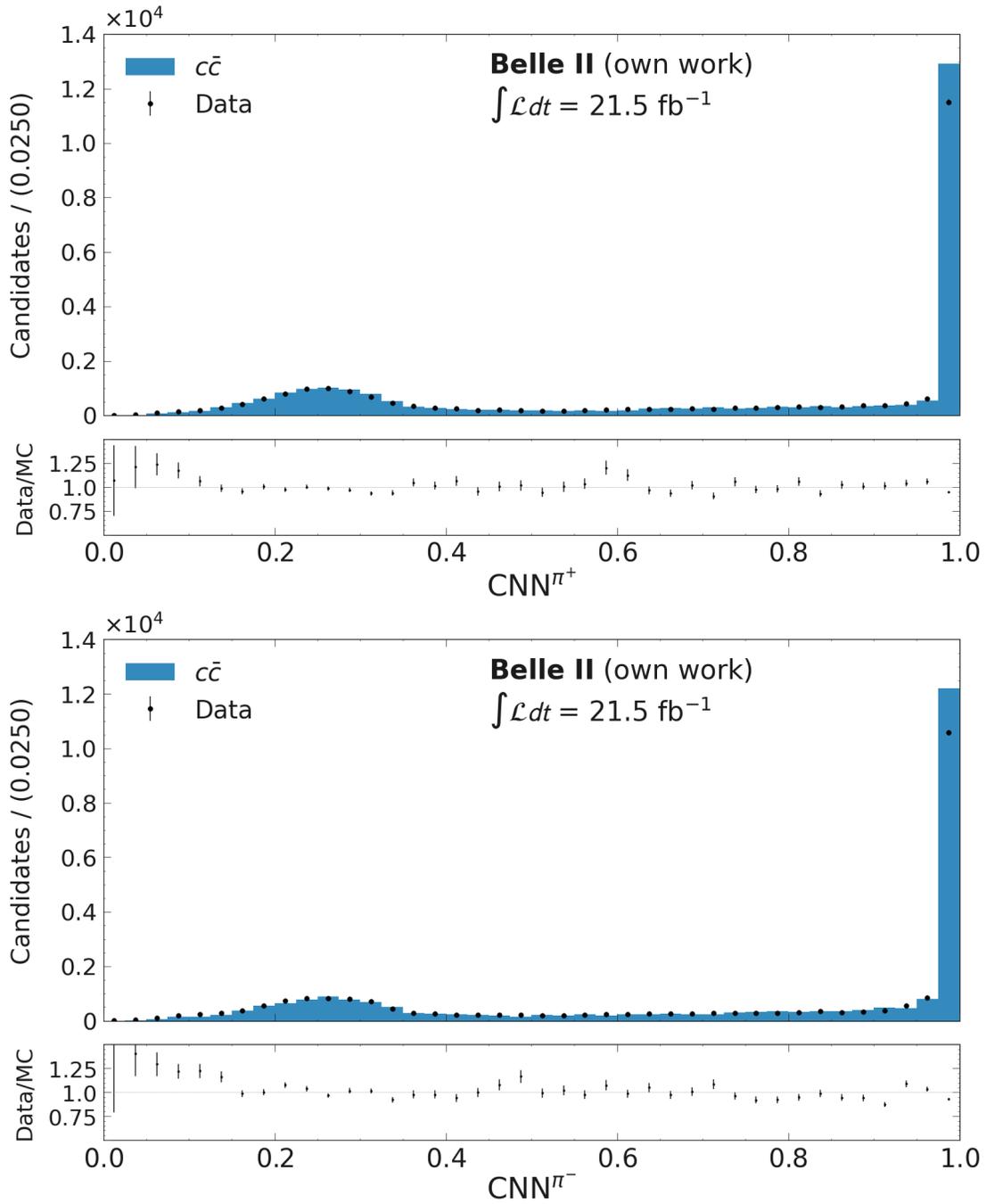


Figure 6.11: Distribution of CNN outputs for π^+ (top) and π^- (bottom) from the D^* channel. The MC signal is normalised to the integrated luminosity of 21.5 fb^{-1} .

6.6 Comparison of CNN and Other PID Methods

Quantitative assessments of the CNN’s performance are made through the computation of the AUC of the ROC for each particle type and dataset. The AUC values provide a measure of the CNN’s capability in distinguishing the signal within the data and MC. A close match in the AUC values between data and MC indicates a successful modeling of the signal by the CNN.

6.6.1 ROC Curve Analysis

The ROC curves depicted in Figures [6.12](#) and [6.13](#) provide a visual representation of the performance of different PID methods across the specified transverse momentum p_T ranges. These curves are crucial in evaluating the trade-off between signal efficiency (true positive rate) and background rejection ($1 - \text{false positive rate}$) at various threshold settings.

For the comprehensive range of $0.28 \leq p_T < 1.0 \text{ GeV}/c$, the ROC curves demonstrate that the CNN delivers a robust performance, indicated by the high AUC values for both data and MC. This consistency across the data and MC suggests that the CNN has effectively learned patterns that are not merely artifacts of the MC but are intrinsic to the physical processes being observed. The CNN’s AUC values do not exhibit a significant overfitting to the MC, indicating its potential for generalization to unseen data.

The BDT and default methods display variances between data and MC AUC values that are similar to those of the CNN. This indicates a comparable level of performance among the methods when it comes to generalizing from MC to real data. Despite the different methodologies and complexity levels inherent to each approach, their ability to capture and utilize discriminative features appears to be effectively consistent. However, this assessment is based solely on AUC comparisons. It’s possible that other metrics or deeper analyses might reveal nuanced differences in performance, such as varying sensitivities to certain data characteristics or potential overfitting to the MC. These aspects warrant further investigation to ensure the robustness of the simulation models and the reliability of the methods across various data sets.

The ROC curves in the segmented momentum ranges-particularly notable in the middle segment $0.52 \leq p_T < 0.76 \text{ GeV}/c$ -highlight the fine-grained performance differences between the PID methods. In these more narrowly defined ranges, the CNN’s adaptability to variations in particle behavior with momentum is evident. The performance of the CNN in these segmented ranges is crucial for its applicability to a broad spectrum of experimental conditions where the momentum of the particles can greatly influence the identification efficiency. The observed results from the ROC analysis suggest that employing a CNN could lead to more accurate PID.

6.6.2 Efficiency and Fake Rate Comparison

The comparison of muon efficiencies at fixed pion fake rates, presented in Tables [6.4](#), [6.6](#), [6.8](#), and [6.10](#), indicates the CNN’s ability to retain a higher signal identification rate at lower

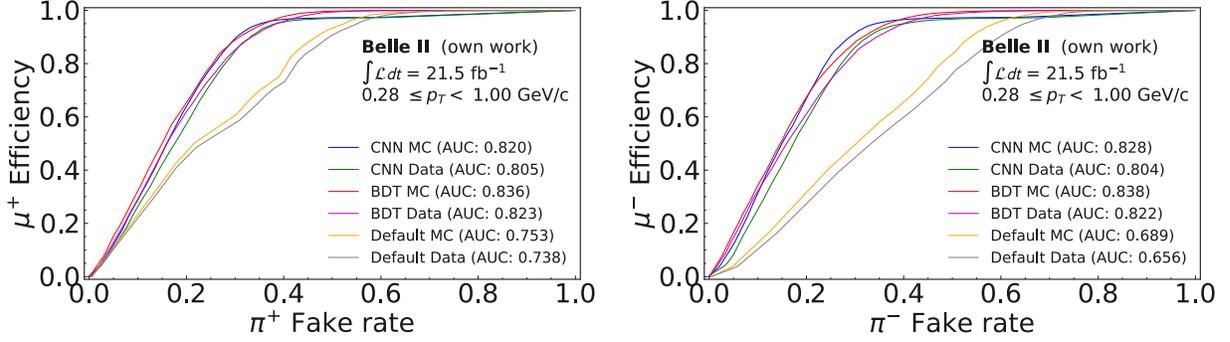


Figure 6.12: ROC curves showing the discriminative performance of various PID methods for data and MC samples in the transverse momentum range $0.28 \leq p_T < 1.00 \text{ GeV}/c$.

levels of background contamination. This is crucial for analyses where the purity of the signal is paramount. The CNN outperforms the default method and is competitive with the BDT, showing its efficacy in signal-background separation.

Similarly, the comparison of pion fake rates at fixed muon efficiencies, as detailed in Tables [6.5](#), [6.7](#), [6.9](#), and [6.11](#), provides insights into the CNN’s performance in maintaining a low rate of mis-identification across various transverse momentum ranges. Consistency in fake rates between data and MC is indicative of the robustness of the CNN model. In-depth analysis of the performance tables for the range of $0.28 \leq p_T < 1.0 \text{ GeV}/c$ reveals several key points. Focusing on the intermediate momentum range $0.52 \leq p_T < 0.76 \text{ GeV}/c$, as documented in Tables [6.8](#) and [6.9](#), the CNN’s performance is noteworthy.

For the given range, the CNN significantly outperforms the default method with a pion fake rate of 10 %, boasting a muon efficiency of 33.28 % for positively charged particles and 37.17 % for negatively charged particles in MC. The CNN performance competes well with the BDT performance, confirming its promise as a sophisticated PID tool.

At a fixed muon efficiency of 80 %, the CNN demonstrates a remarkable ability to minimize the pion fake rate to 27.07 % for positively charged and 26.47 % for negatively charged particles in MC samples. These figures are notably lower than those achieved by the default method, and they often surpass the BDT, which highlights the CNN’s utility in reducing background noise. This capability is especially important in the intermediate momentum range, where detection is optimal, and effects such as multiple scattering are minimized.

The slight differences between data and MC efficiencies and fake rates for the CNN are indicative of a model that is well-tuned and not overfitting to the MC, suggesting that it is capturing the true underlying distributions.

Overall, the CNN’s robust performance across various momentum ranges and particle charges, and the close agreement between data and MC results, reinforce its viability as a reliable and effective tool for PID in Belle II.

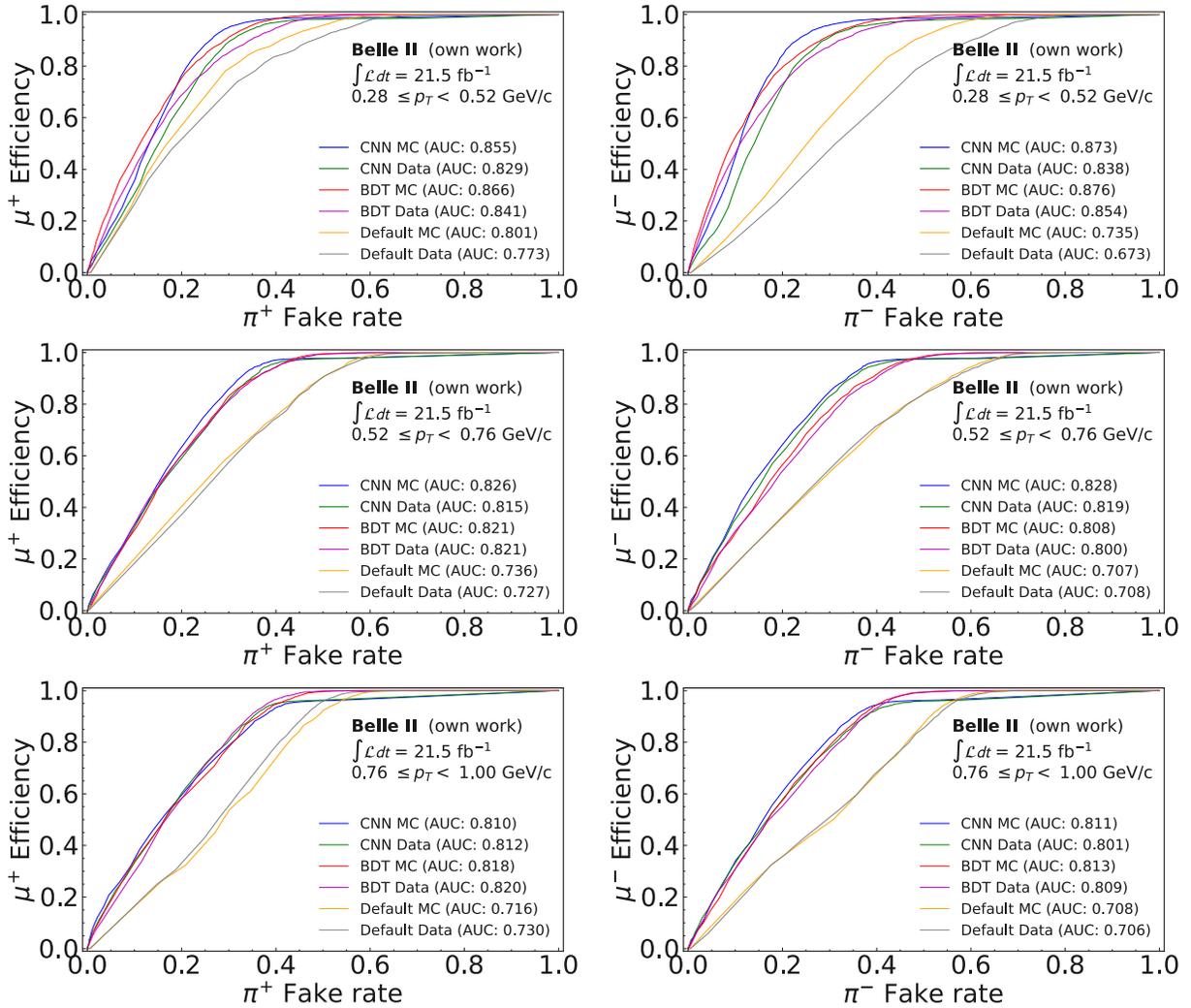


Figure 6.13: ROC curves showing the discriminative performance of various PID methods for data and MC samples in three transverse momentum ranges between 0.28 and 1.0 GeV/c.

Table 6.4: Comparison of μ efficiencies at fixed π fake rates for various PID methods in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c .

Charge	Fake Rate (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	10	MC	28.97	32.18	23.24
		Data	26.29	28.81	21.96
	20	MC	63.95	65.35	46.86
		Data	57.33	61.93	44.80
	30	MC	89.43	88.82	60.54
		Data	84.67	85.14	57.69
-	10	MC	31.33	34.24	12.70
		Data	25.21	32.10	10.30
	20	MC	66.95	67.38	31.58
		Data	58.55	61.09	27.39
	30	MC	91.38	88.05	49.55
		Data	86.12	84.04	44.73

Table 6.5: Comparison of π fake rates at fixed μ efficiencies for various PID methods in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c .

Charge	Efficiency (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	80	MC	25.73	25.38	40.65
		Data	27.96	27.36	42.63
	90	MC	30.32	30.89	46.61
		Data	33.13	33.36	49.37
	95	MC	35.16	35.98	52.39
		Data	38.30	38.54	55.10
-	80	MC	23.98	25.37	47.71
		Data	27.19	27.98	52.46
	90	MC	28.92	31.64	53.75
		Data	32.65	34.47	59.68
	95	MC	34.27	36.73	58.66
		Data	40.05	39.99	64.77

Table 6.6: Comparison of μ efficiencies at fixed π fake rates for various PID methods in the transverse momentum range $0.28 \leq p_T < 0.52$ GeV/ c .

Charge	Fake Rate (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	10	MC	35.63	45.93	28.21
		Data	30.66	40.17	26.43
	20	MC	76.46	75.40	56.97
		Data	64.95	68.79	51.72
	30	MC	95.04	91.19	78.98
		Data	89.43	85.23	70.57
-	10	MC	44.86	52.51	17.07
		Data	33.12	46.13	13.00
	20	MC	84.27	79.54	38.04
		Data	72.17	72.96	29.35
	30	MC	95.91	91.82	60.11
		Data	91.18	87.58	47.66

Table 6.7: Comparison of π fake rates at fixed μ efficiencies for various PID methods in the transverse momentum range $0.28 \leq p_T < 0.52$ GeV/ c .

Charge	Efficiency (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	80	MC	21.19	22.24	30.87
		Data	25.15	26.28	36.76
	90	MC	26.07	29.19	40.98
		Data	30.50	34.48	47.86
	95	MC	29.96	34.00	48.57
		Data	35.79	40.47	55.69
-	80	MC	14.20	16.88	47.29
		Data	18.91	21.20	53.39
	90	MC	20.18	25.50	57.12
		Data	25.47	30.77	63.81
	95	MC	24.51	31.75	64.67
		Data	30.71	37.69	71.52

Table 6.8: Comparison of μ efficiencies at fixed π fake rates for various PID methods in the transverse momentum range $0.52 \leq p_T < 0.76$ GeV/ c .

Charge	Fake Rate (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	10	MC	33.28	31.17	20.14
		Data	32.24	33.41	18.37
	20	MC	63.36	60.32	40.33
		Data	59.07	60.29	37.20
	30	MC	85.88	83.12	59.60
		Data	81.84	81.32	57.43
-	10	MC	37.17	29.98	18.08
		Data	35.35	30.85	17.88
	20	MC	64.32	56.49	35.98
		Data	61.42	54.08	36.44
	30	MC	84.84	77.72	53.69
		Data	83.05	75.10	54.83

Table 6.9: Comparison of π fake rates at fixed μ efficiencies for various PID methods in the transverse momentum range $0.52 \leq p_T < 0.76$ GeV/ c .

Charge	Efficiency (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	80	MC	27.07	28.76	42.99
		Data	29.14	29.26	43.29
	90	MC	32.26	35.30	49.72
		Data	34.70	35.70	49.59
	95	MC	36.65	40.60	54.61
		Data	38.64	40.74	55.11
-	80	MC	26.47	29.85	40.76
		Data	28.19	30.19	41.90
	90	MC	31.59	36.40	48.30
		Data	33.98	36.76	48.01
	95	MC	35.95	41.50	53.16
		Data	37.40	42.10	54.25

Table 6.10: Comparison of μ efficiencies at fixed π fake rates for various PID methods in the transverse momentum range $0.76 \leq p_T < 1.0$ GeV/ c .

Charge	Fake Rate (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	10	MC	34.72	33.99	16.30
		Data	33.27	29.78	16.58
	20	MC	59.51	58.19	31.32
		Data	60.24	58.12	33.15
	30	MC	78.68	78.19	53.51
		Data	80.50	81.56	55.38
-	10	MC	33.87	30.90	18.57
		Data	33.26	30.86	17.20
	20	MC	60.82	57.43	35.71
		Data	57.76	55.23	35.67
	30	MC	81.70	78.67	49.01
		Data	77.79	76.29	50.85

Table 6.11: Comparison of π fake rates at fixed μ efficiencies for various PID methods in the transverse momentum range $0.76 \leq p_T < 1.0$ GeV/ c .

Charge	Efficiency (%)	Sample Type	CNN (%)	BDT (%)	Default (%)
+	80	MC	30.77	30.93	42.72
		Data	29.77	29.35	40.73
	90	MC	37.26	36.29	48.76
		Data	35.19	34.52	46.05
	95	MC	42.56	41.79	54.31
		Data	40.02	38.89	52.14
-	80	MC	31.11	30.02	43.99
		Data	30.89	29.54	42.21
	90	MC	38.66	37.59	50.87
		Data	37.14	36.77	49.11
	95	MC	43.55	42.98	56.22
		Data	41.89	40.25	54.55

6.7 CNN Evaluation on Energy Thresholds

The evaluation of CNN performance on different energy thresholds is critical to verify the method’s robustness in distinguishing between signal and background. The analysis is based on the ROC curves generated from testing data that passed various energy thresholds to mitigate background impact, as depicted in Figure 6.14 and 6.15. These thresholds, applied post-training during the testing phase, serve as a filter to enhance signal purity by suppressing potential background contributions in data. To more accurately discern the differences, the axis limits are deliberately constrained, thereby sharpening the visualization of subtle variations in performance.

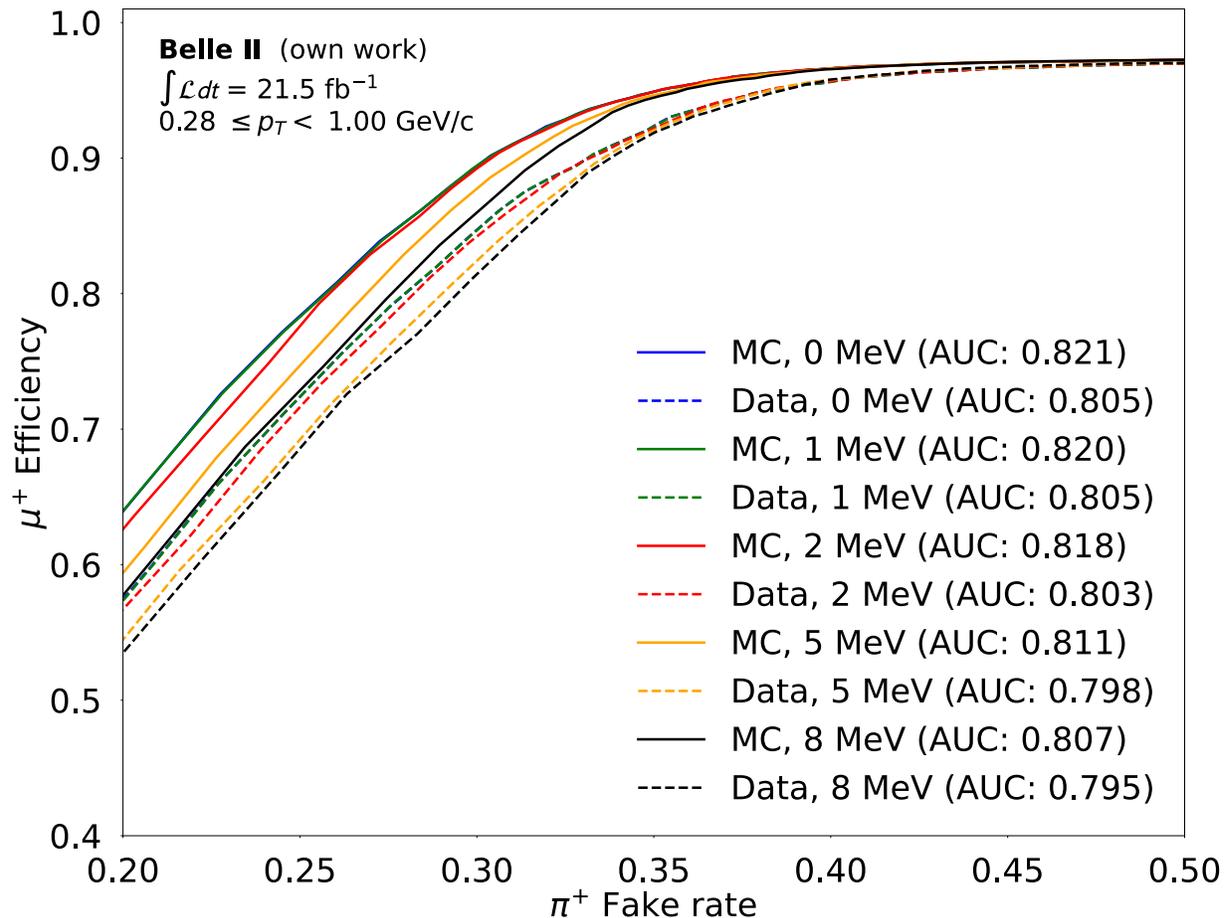


Figure 6.14: ROC curves showing the discriminative performance of various energy thresholds for data and MC samples in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for positive charged muons and pions.

With the negative charged tracks, the CNN exhibits a robust initial classification power with high AUC values for both the MC and data. As the threshold increases, a slight reduction in AUC values is observed, suggesting a systematic bias in the PID performance introduced by the trade-off between background suppression and potential loss of low-energy signal events. The drop in discrimination power with increasing energy thresholds,

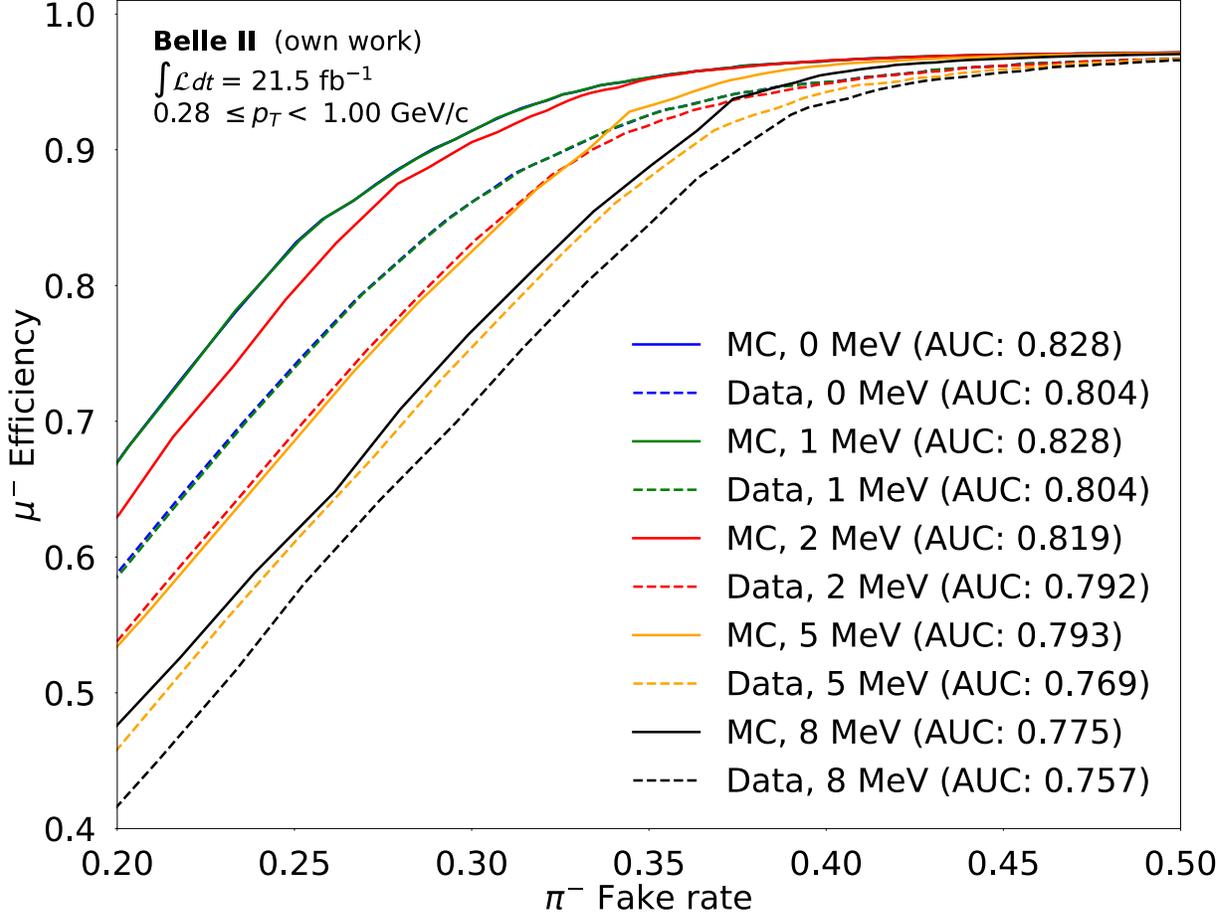


Figure 6.15: ROC curves showing the discriminative performance of various energy thresholds for data and MC samples in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for negative charged muons and pions.

particularly at 5 MeV and 8 MeV, may reflect the exclusion of subtle yet significant signal features, necessitating a careful balance to avoid systematic underestimation of PID capabilities.

Conversely, positive charged tracks demonstrate a more consistent AUC across MC and data, indicating the CNN’s resilience to the exclusion of lower-energy signals up to a certain threshold. The gradual performance decrease beyond this point requires an assessment of the systematic uncertainties that such thresholding introduces, particularly in light of the apparent charge asymmetry. This asymmetry is evident in the ROC curve analysis, where negatively charged particles show a greater reduction in AUC compared to positively charged particles, suggesting the need to quantify and correct for charge-dependent systematic effects.

The close correspondence between AUC values for MC and data highlights the CNN’s generalization capabilities. However, any emergent differences with threshold variations must be quantified as systematic uncertainties, especially since the simulation may not perfectly model the impact of energy thresholds. Additionally, the application of energy

thresholds affects energy calibration and resolution, which could systematically shift the measured energy spectrum and introduce a further source of systematic error if this shift differs between MC and data.

In summary, the energy thresholds used in CNN-based PID not only act as a tool for enhancing signal-to-noise ratio but also introduce additional layers of complexity that must be considered for systematic effects. The observed charge-dependent discrepancies and the potential for information loss through excessive thresholds necessitate a judicious application of energy thresholds. To ensure accuracy in experimental analyses, it is crucial to balance noise reduction and signal retention while accounting for systematic variations introduced by thresholds, as reflected in the systematic effects analysis and the ROC curve diagnostics.

6.8 CNN Evaluation with Various Beam Backgrounds

In the context of PID using CNNs, the robustness and adaptability of the network are scrutinized under varying conditions of beam-induced background. The ROC curves, as depicted in Figure [6.16](#) and [6.17](#), are analyzed for different levels of beam-induced background. Specifically, the impact of single $BG \times 1$, double $BG \times 2$, and quintuple $BG \times 5$ nominal beam backgrounds on the PID effectiveness is evaluated for both positive and negative muons and charged pions across MC and data.

The PID capabilities of the CNN are observed to be differentially affected by the beam background conditions. The AUC values for both MC and data indicate that the CNN retains a high level of performance under enhanced beam backgrounds, yet nuances in the response are evident. It is discerned that the CNN's ability to classify negatively charged muons and pions exhibits a greater sensitivity to the increase in beam backgrounds compared to positively charged muons and pions. This discrepancy in the response is highlighted by a more pronounced variation in the AUC for negatively charged muons and pions as the background conditions intensify.

In the analysis of positively charged particles, the ROC curves reveal a more stable discriminative capability of the CNN across the various beam background scenarios. The AUCs for MC and data remain comparatively consistent, suggesting that the CNN model, trained with nominal beam background conditions, generalizes well to elevated background conditions for positively charged muons and pions.

The differences in AUC between data and MC, and between positively and negatively charged muons and pions, are accentuated when the axis limits are constrained to enhance visualization. This approach allows for a more detailed observation of the subtle variations in performance that might otherwise be overshadowed by the dominant signal regions of the ROC space.

The disparities observed call attention to potential systematic effects that could arise from varying beam backgrounds. Such effects are crucial to consider when extrapolating findings from MC to data, particularly in experimental settings where the beam conditions

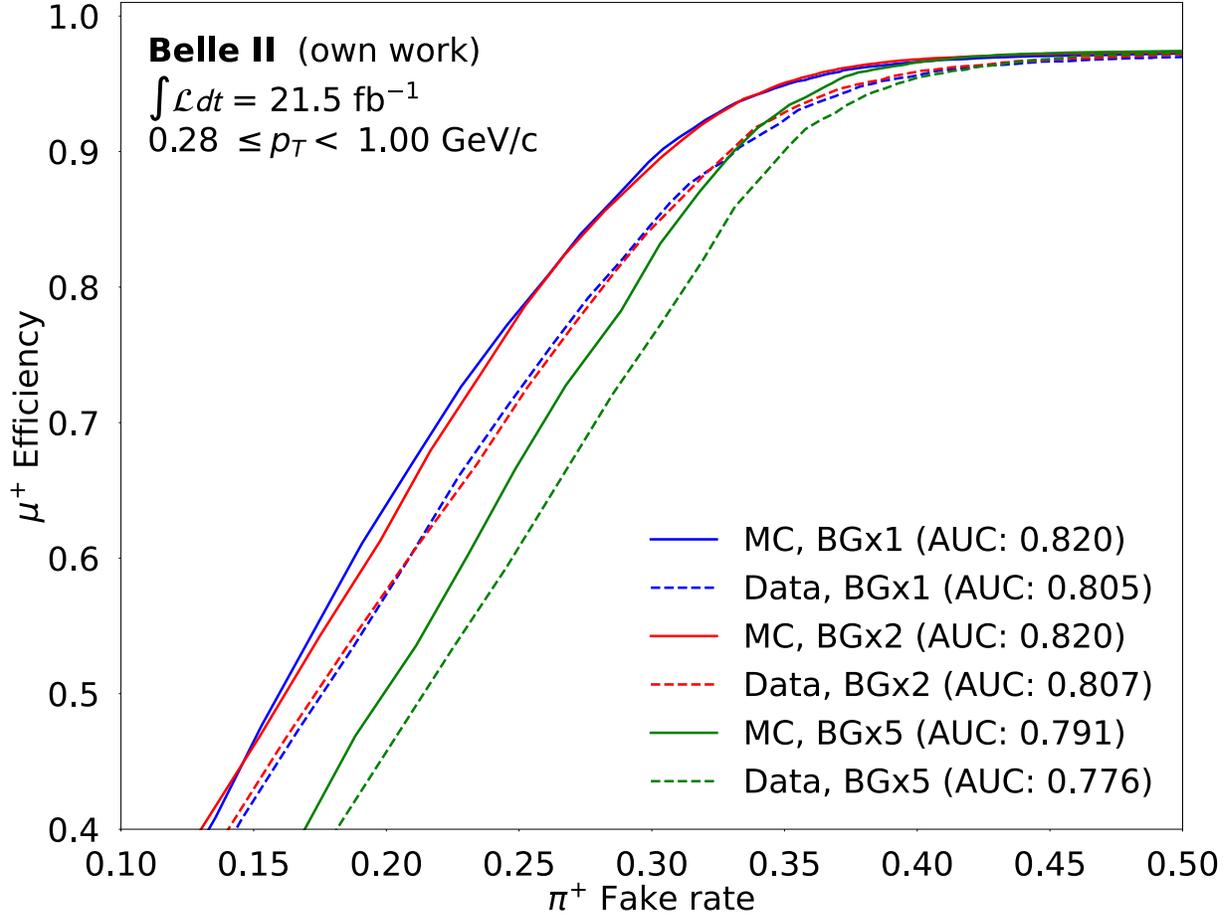


Figure 6.16: ROC curves showing the discriminative performance of various beam backgrounds for data and MC samples in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for positive charged muons and pions.

can fluctuate. The findings highlight the importance of incorporating beam background variations into the systematic uncertainty estimations for experiments relying on CNN-based PID methods.

In summary, the CNN demonstrates a resilient PID performance against increasing beam backgrounds, particularly for positively charged particles. However, the variations observed for negatively charged particles and between data and MC samples necessitate a careful consideration of systematic uncertainties to ensure accurate and reliable particle identification in environments with fluctuating beam conditions.

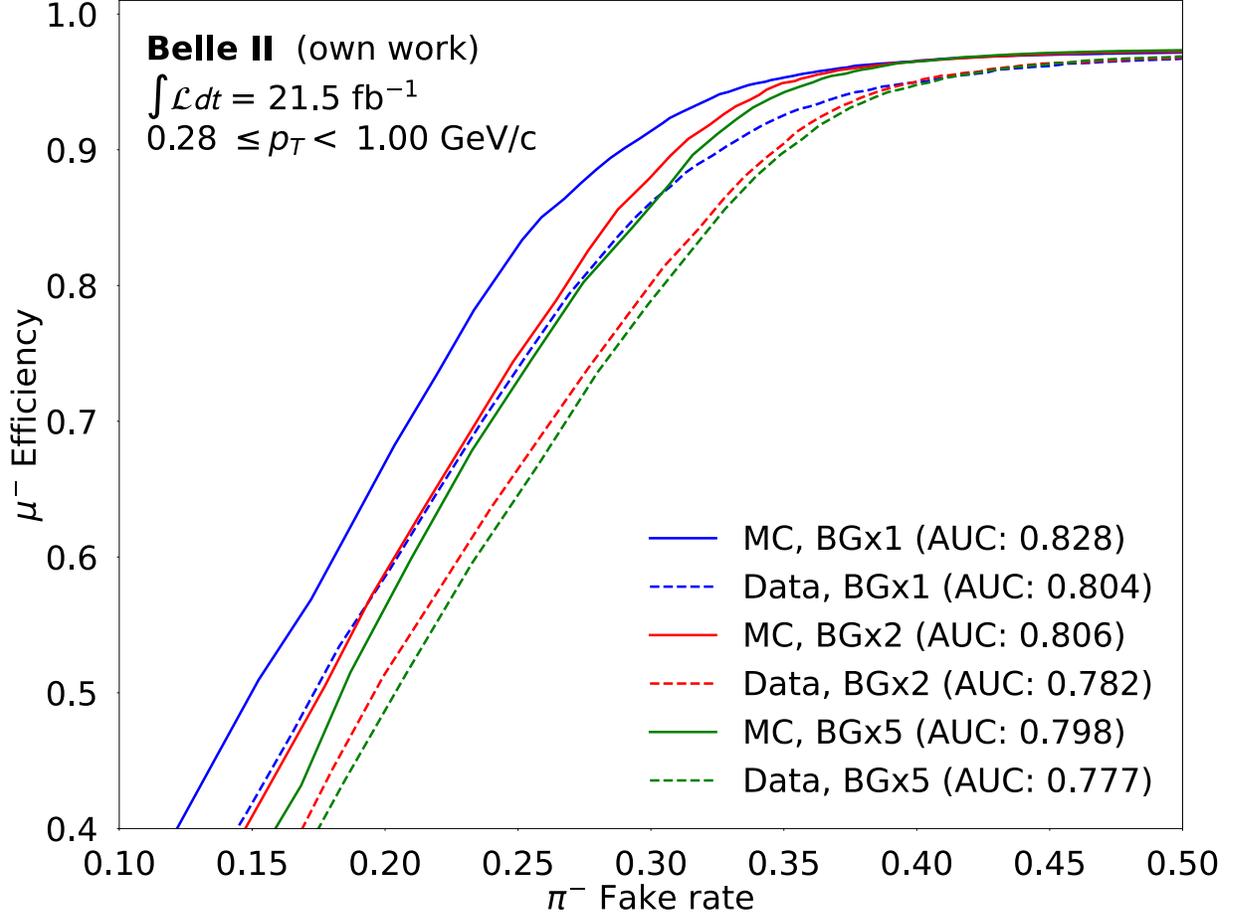


Figure 6.17: ROC curves showing the discriminative performance of various beam backgrounds for data and MC samples in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for negative charged muons and pions.

6.9 Isolation Effects

The isolation criteria significantly influence the PID capabilities of CNNs. This section focuses on the isolation effects, particularly focusing on the CNN output’s distinctive peak structure observed for muons. Unlike pions, muons exhibit a less dispersed structure in the pixel images representing energy depositions. This distinction becomes blurred when another particle, such as a muon, is in close proximity to the particle under observation, leading to a more dispersed pattern in the energy deposition images.

The isolation degree of a particle, defined by its closeness to the nearest track or photon, crucially impacts the CNN’s accuracy in particle classification. This section investigates how particle isolation affects the PID performance of CNNs.

6.9.1 Mis-classification Analysis of Muon Tracks

Initially, the study examines pixel images of muon tracks that the CNN has mis-classified with probabilities below 0.05. These images, displayed in Figure 6.18, show instances where muons are mistakenly identified as pions. The analysis seeks to identify common feature contributing to these mis-classifications. Notably, high-energy pixels, predominantly exceeding 1 GeV, are observed near the tracks, indicating activity unrelated to the muons themselves.

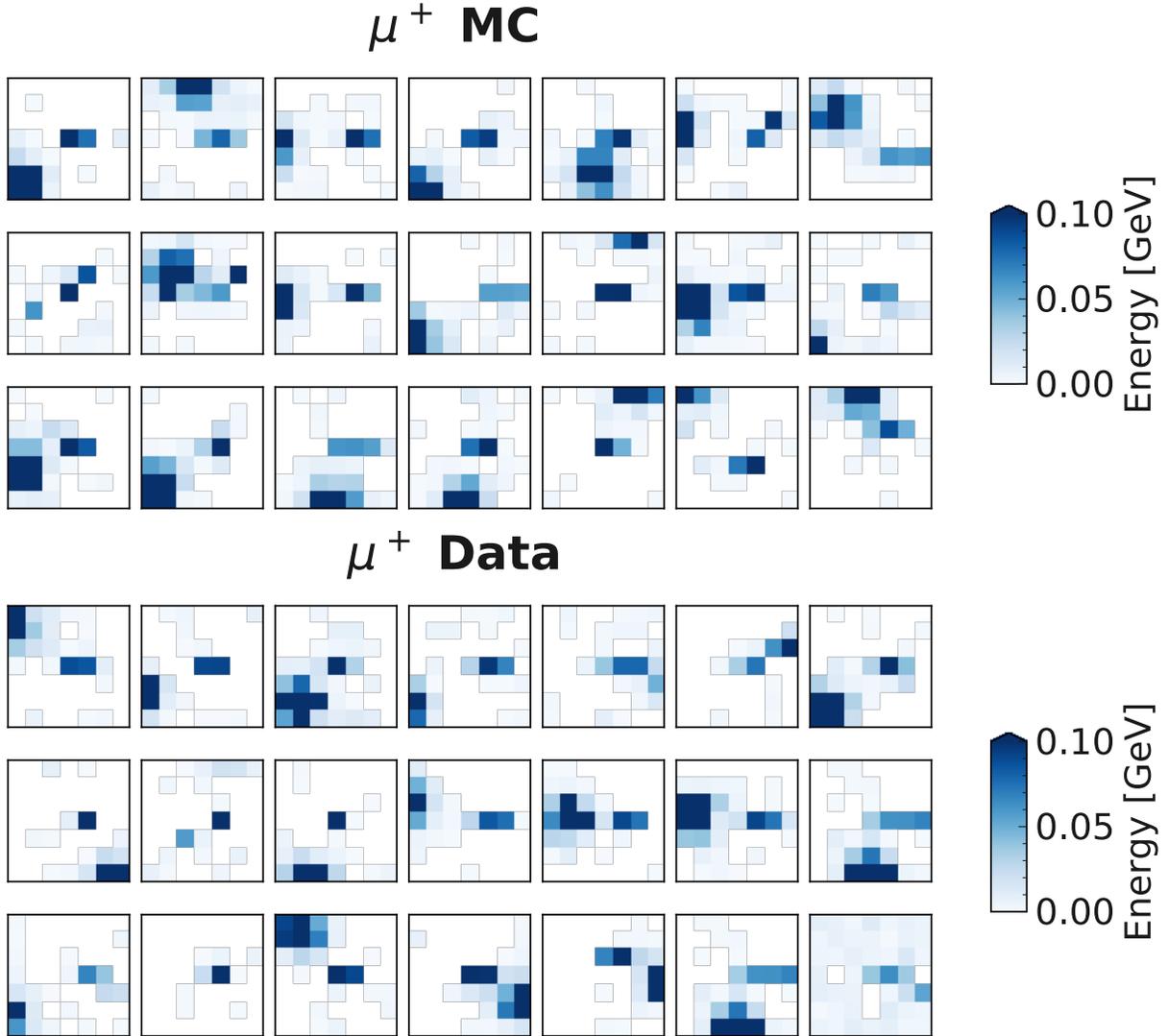


Figure 6.18: Muons' energy depositions in pixel images for MC (top) and data (bottom) when CNN output of muon is less than 0.05.

In this context, factors like overlapping tracks and proximity to photons or clusters are scrutinized. For both positive and negative muons, distances to the nearest track and photon (or cluster) are computed for cases with CNN probabilities below and above 0.05. These measurements serve as indicators of particle isolation within the detector,

potentially influencing CNN classification performance. While the distance distributions to the nearest tracks do not yield significant insights, those related to the nearest photon reveal a pronounced peak at approximately 25 cm, equivalent to the span of 5 crystals. This finding corroborates the observation of high-energy pixels in Figure 6.18.

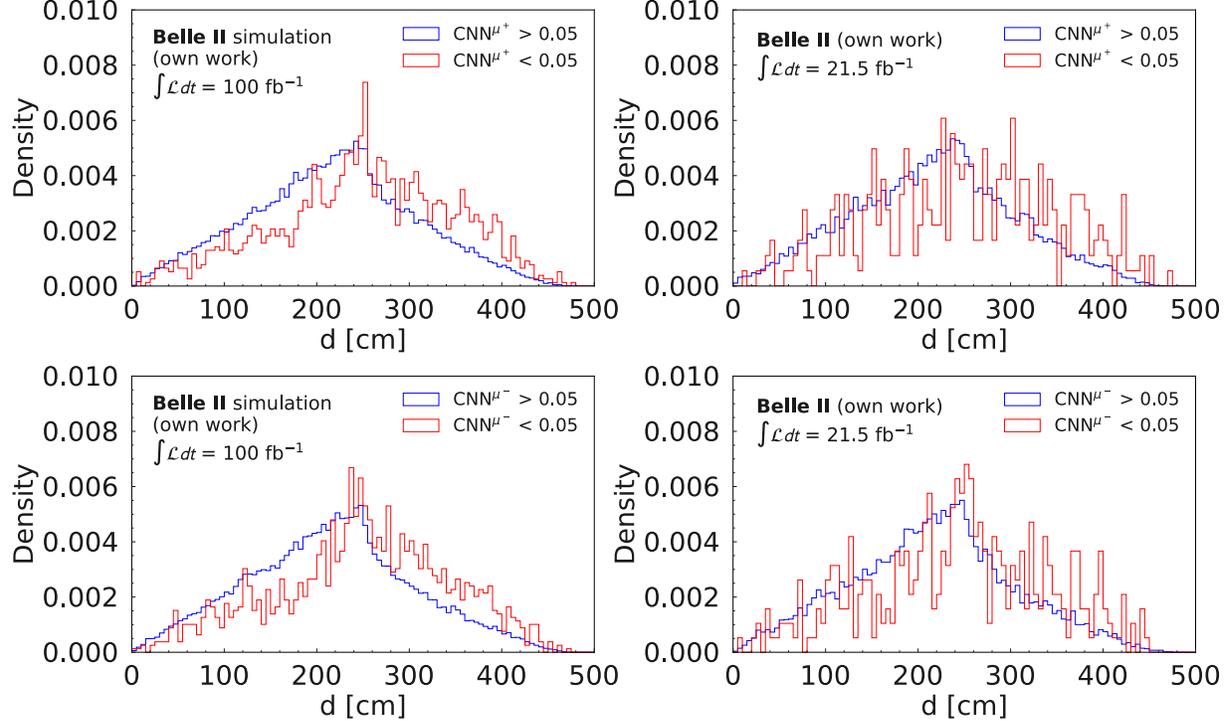


Figure 6.19: Distribution of distances of muons to the nearest track for μ^+ (top) and μ^- (bottom) for MC (left) and data (right).

6.9.2 Evaluating the Impact of Isolation Cuts

To enhance the CNN’s PID efficacy, an isolation cut is proposed. This approach entails excluding muon tracks within 50 cm of a photon from the analysis. The hypothesis is that reducing the presence of nearby photons will improve the CNN’s distinction between muons and charged pions. Figure 6.21 shows the CNN output distributions following isolation cut for both types of muons, showing a marked reduction in mis-identified cases.

6.9.3 Improving CNN Performance with Isolation Cuts

The effectiveness of this isolation cut is further validated by recalculating the ROC curve and comparing it to the baseline CNN method. This comparison, illustrated in Figure 6.22, encompasses positive and negative charged tracks for all and isolated muons across various p_T ranges. The analysis is segmented into four p_T categories:

- all: $0.28 \leq p_T < 1.0$ GeV/c,

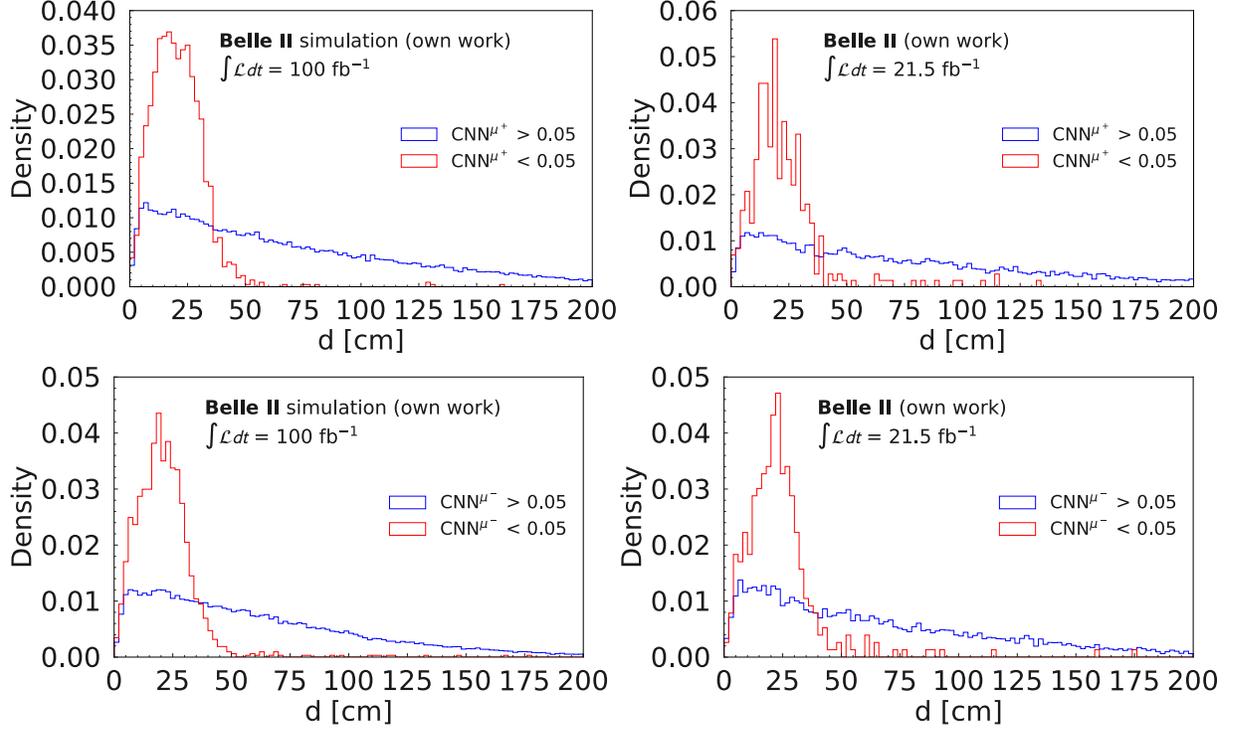


Figure 6.20: Distribution of distances of muons to the nearest photon for μ^+ (top) and μ^- (bottom) for MC (left) and data (right).

- low: $0.28 \leq p_T < 0.52$ GeV/ c ,
- intermediate: $0.52 \leq p_T < 0.76$ GeV/ c ,
- high: $0.76 \leq p_T < 1.0$ GeV/ c ,

facilitating a deeper understanding of the isolation cut's impact across different momentum spectrums. Comparing these ROC curves with the baseline (without isolation cuts) highlights the effectiveness of the isolation strategy in enhancing PID accuracy. These ROC curves show a clear improvement in the CNN's PID performance, particularly in higher p_T ranges. The ROC curves signify enhanced sensitivity and specificity in identifying muons accurately.

The findings from this analysis highlight the critical role of isolation in the PID process using CNNs. Isolation, particularly from photons, emerges as a key factor in reducing mis-classifications. The introduction of an isolation cut, based on the proximity to photons, proves to be a highly effective strategy in enhancing the CNN's accuracy for muon identification.

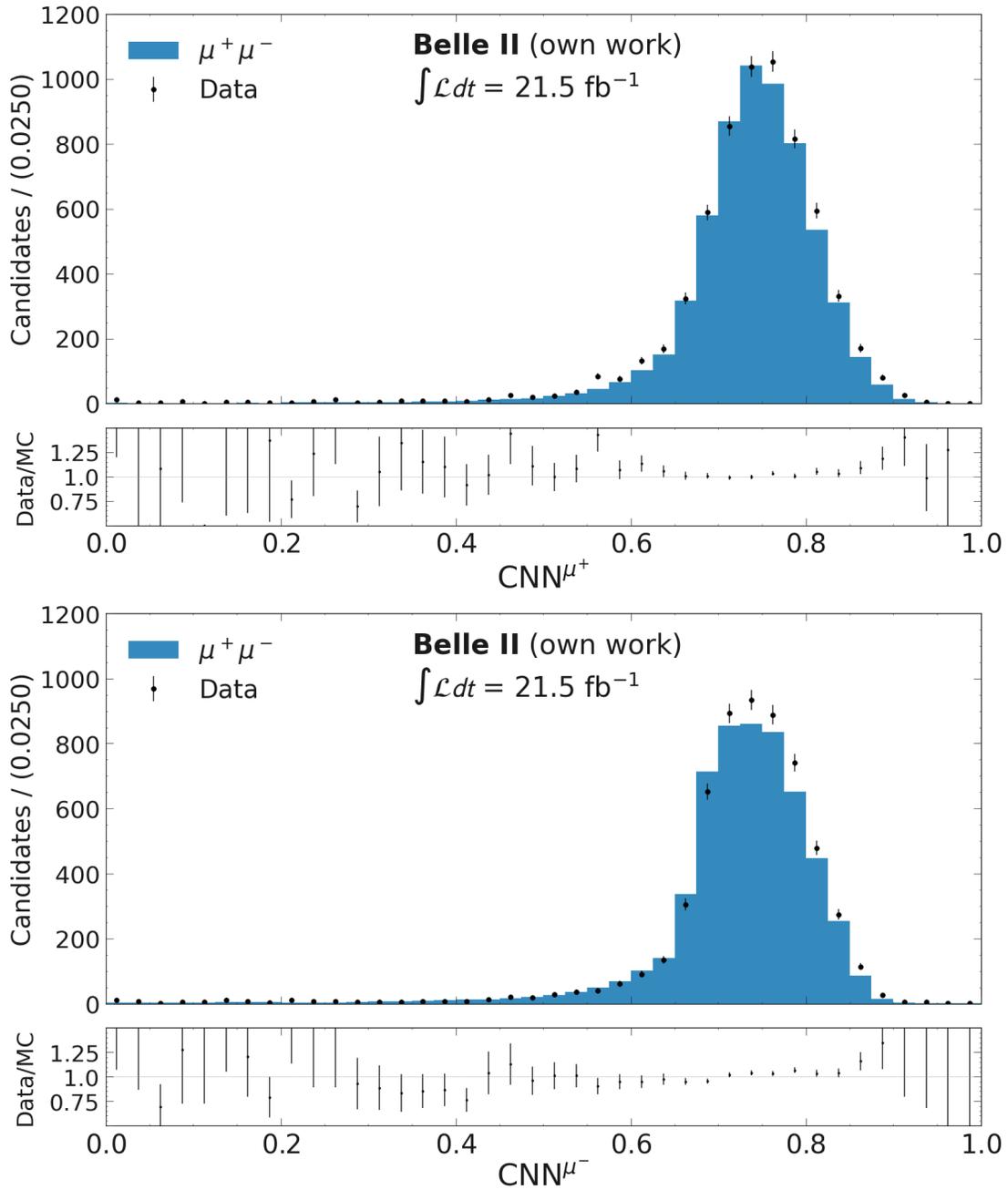


Figure 6.21: Distribution of CNN outputs for μ^+ (top) and μ^- (bottom) from the radiative dimuon channel after excluding muon tracks with a nearby photon.

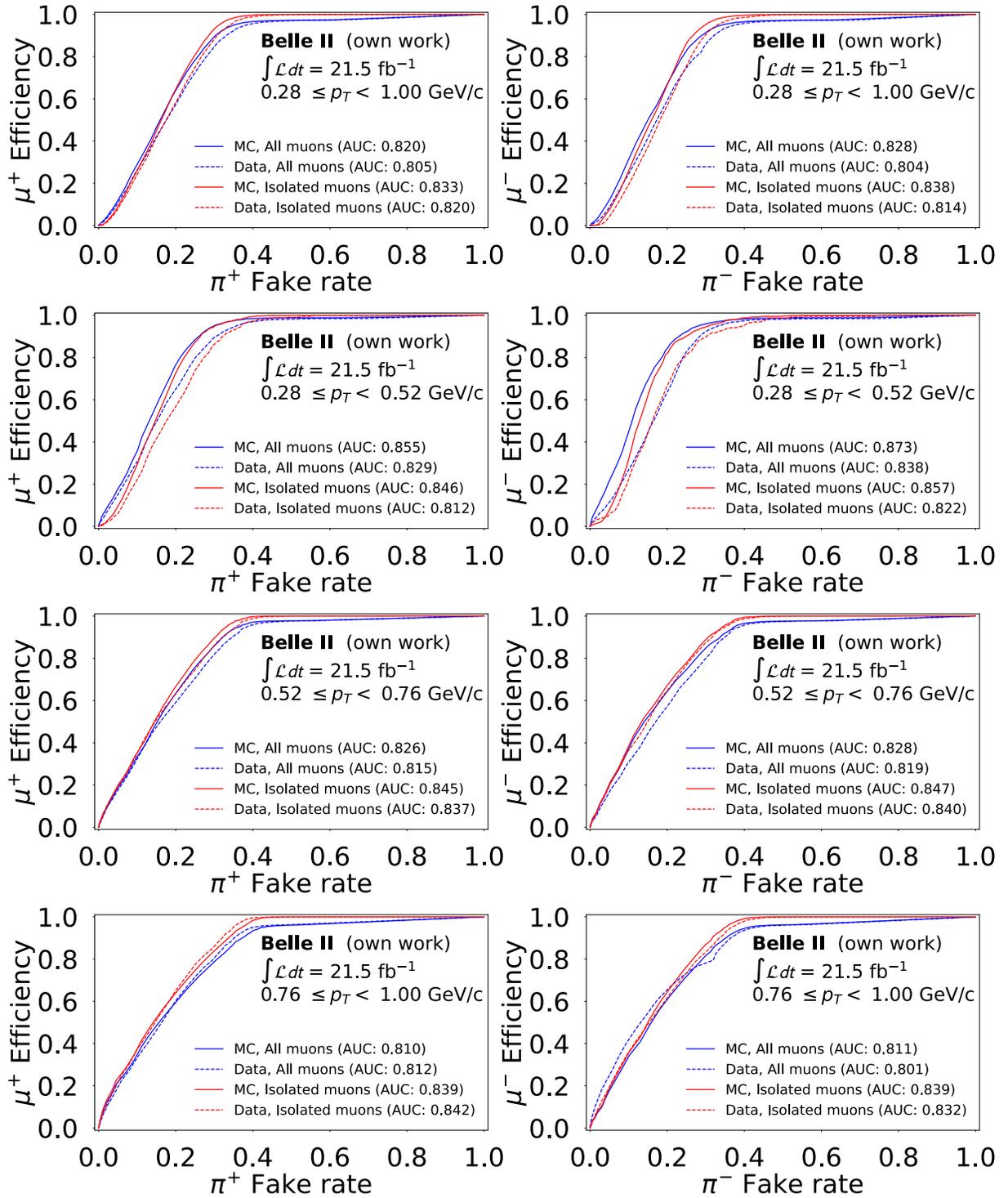


Figure 6.22: ROC curves for positive (left) and negative (right) charged tracks for all and isolated muons across different p_T ranges for CNN method. From top to bottom: all ($0.28 \leq p_T < 1.0 \text{ GeV}/c$), low ($0.28 \leq p_T < 0.52 \text{ GeV}/c$), intermediate ($0.52 \leq p_T < 0.76 \text{ GeV}/c$), and high ($0.76 \leq p_T < 1.0 \text{ GeV}/c$).

Chapter 7

Summary and Outlook

The research detailed in this dissertation demonstrates that employing patterns of energy depositions in the Belle II Electromagnetic Calorimeter (ECL) enhances muons and charged pions separation, particularly for tracks with low momentum. The study utilized a particle gun to simulate single-track events featuring both muons and charged pions in the ECL barrel. Considering the unique geometry of the ECL, two distinct Convolutional Neural Networks (CNNs) were trained, each trained specifically for positively and negatively charged tracks.

The CNN models' effectiveness was evaluated using a test dataset generated under conditions identical to the training dataset. The results showed that the CNN Particle Identification (PID) method outperforms the existing Boosted Decision Trees (BDT) PID and standard binary PID methods used by Belle II, particularly in low transverse momentum regions. Notably, in the intermediate transverse momentum range ($0.52 \leq p_T < 0.76$ GeV/ c), the CNN PID method surpassed the BDT PID method by 2.60 % and 4.46 % for positively and negatively charged tracks, respectively. In the lower range ($0.28 \leq p_T < 0.52$ GeV/ c), the improvement over standard binary PID method was even more significant: 17.95 % and 31.24 % for positive and negative tracks, respectively. Furthermore, a comparison across the entire transverse momentum range ($0.28 \leq p_T < 1.0$ GeV/ c) revealed an improvement of 3.51 % and 6.97 % for positive and negative tracks, respectively, when comparing CNN and BDT PID methods at a fixed pion fake rate of 20 %. At a fixed muon efficiency of 99 %, the improvement in pion fake rate was 5.70 % and 9.93 % for positive and negative tracks, respectively. A key advantage of the CNN PID method is its independence from clustering algorithms, unlike the BDT PID method, which relies heavily on shower shape information, and the standard binary PID method, which uses likelihoods.

The study further explored the performance of the CNN models under various energy threshold conditions, which is crucial for mitigating noise in pixels during high beam background scenarios. The relationship between the energy threshold and model performance is complex and non-linear. Higher thresholds, while reducing noise, can also inadvertently filter out critical data, potentially impairing the model's performance.

Additionally, the CNN models were trained and tested on samples generated with different beam backgrounds, originating from interactions between the primary beams

and various components of the machine. The findings underscored that while training a model with a specific background can enhance performance in that context, it does not necessarily ensure optimal performance across diverse background conditions. Therefore, it's essential to comprehend how different beam backgrounds interact with the model's training to maximize performance. The study of different conditions for beam backgrounds and thresholds can be translated as systematic uncertainties in the future.

In real-world application, the CNN PID method was tested using an on-resonance collision dataset with an integrated luminosity of $\int L dt = 21.5 \text{ fb}^{-1}$, alongside Monte Carlo (MC) samples. For validation, clean muon and pion samples were selected using $e^+e^- \rightarrow \mu^+\mu^-\gamma$ and $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$, respectively. These samples were examined by investigating the mass distributions of both data and MC signals to ensure accurate event selection. The trained CNN models, initially calibrated with single-track events generated via particle gun, were then tested on the MC and collision data containing multi-track events. To underscore the superiority of the CNN PID method over other PID methods, the study meticulously calculated both muon efficiencies at fixed pion fake rates and pion fake rates at fixed muon efficiencies. In comparisons, the CNN PID method showed remarkable improvements over the BDT PID method, with enhancements reaching up to 8% for both muon efficiencies and pion fake rates. Even more striking, when compared with the standard binary PID method, the CNN PID method demonstrated significant improvements — up to 44% in muon efficiencies and 41% in pion fake rates. These improvements were particularly notable in low and intermediate transverse momentum regions, emphasizing the CNN method's effectiveness across a range of scenarios.

The research further explored how particles interact spatially within the detector, taking into account the sensitivity of energy deposition in the crystals to nearby particles. A key finding was that implementing an isolation cut significantly improves the CNN models' performance, notably in reducing the mis-identification of muons. This insight into spatial interactions paves the way for developing algorithms and strategies that enhance the accuracy and efficiency of PID. The study conclusively demonstrates that utilizing isolation effects in PID with CNNs is a powerful approach to refine PID accuracy.

Looking ahead, further enhancements could include integrating low-level ECL crystal information, such as pulse-shape discrimination (PSD) [93, 94] and timing. These features could potentially facilitate the distinction between hadronic and electromagnetic interactions and differentiate between energy depositions from beam backgrounds and actual tracks. There are two potential approaches for integrating these information: one involves introducing these information as an additional layer in the convolution phase of the network, and the other entails training the network initially with energy depositions before incorporating PSD and timing information at a later stage of training.

Given that the current study concentrates exclusively on the ECL barrel, it inherently limits its scope to tracks within this region. Consequently, expanding the analysis to encompass the entire ECL could be highly advantageous for covering a broader phase space. This expansion, however, introduces a challenge: the number of ECL crystals in rings of θ_{ID} outside the barrel do not consistently equal 144, leading to asymmetric pixel images that are not ideally suited for CNN training. Nevertheless, this obstacle could be effectively

navigated using Graph Neural Networks (GNNs). GNNs, with their capability to handle irregular data structures, may offer a promising alternative to address the asymmetry and complexity of the entire ECL.

List of Figures

2.1	The Standard Model of Elementary Particles	5
2.2	Ionization energy loss for various particles in air	10
2.3	Development of an electromagnetic shower	12
3.1	SuperKEKB e^+e^- collider	16
3.2	Belle II top view	18
3.3	Belle II schematic view	18
3.4	Dimension of the barrel crystals	23
3.5	Overall design of the ECL	24
3.6	Polar angle of ECL crystals	26
3.7	Distribution of E/p in low and high momentum ranges	28
4.1	Convolution operation	32
4.2	Typical CNN architecture	34
4.3	Rolling a ball down the valley	38
5.1	Distributions of reconstructed variables generated for events generated under BG \times 1	43
5.2	Typical 7×7 pixel images of muons and pions	44
5.3	Distributions of p_T , θ_{ID} , and ϕ_{ID} for events generated under BG \times 1	45
5.4	Distributions of p_T , θ_{ID} , and ϕ_{ID} in training dataset for events generated under BG \times 1	46
5.5	Distributions of p_T , θ_{ID} , and ϕ_{ID} in validation dataset for events generated under BG \times 1	46
5.6	Distributions of p_T , θ_{ID} , and ϕ_{ID} in test dataset for events generated under BG \times 1	47
5.7	Evolution of model's objective value during hyperparameter tuning	50
5.8	Hyperparameter Significance	51

5.9 CNN architecture.	53
5.10 Training and validation accuracy $BG \times 1$	53
5.11 ROC curve for CNN-based μ/π separation in the ECL	57
5.12 Zoomed ROC curve focusing on low pion fake rates	57
5.13 ROC curve for CNN-based μ/π separation in low p_T region	59
5.14 ROC curve for CNN-based μ/π separation in intermediate p_T region	60
5.15 ROC curve for CNN-based μ/π separation in high p_T region	60
5.16 Comparison of ROC curves for CNN, BDT, and default PID methods over a p_T range of 0.28 to 1 GeV/c	61
5.17 ROC curves for charged tracks across different p_T ranges for different PID methods.	64
5.18 AUC values for positively charged muons and pions using the CNN method across various p_T and θ_{ID} ranges.	66
5.19 AUC values for positively charged muons and pions using the BDT method across various p_T and θ_{ID} ranges.	66
5.20 AUC values for negatively charged muons and pions using the CNN method across various p_T and θ_{ID} ranges.	67
5.21 AUC values for negatively charged muons and pions using the BDT method across various p_T and θ_{ID} ranges.	67
5.22 Muon's 7×7 pixel images at various thresholds	68
5.23 Pion's 7×7 pixel images at various thresholds	68
5.24 ROC Curves for Different Energy Thresholds	69
5.25 Zoomed ROC curve focusing on low π^+ fake rates for various energy thresholds	69
5.26 Zoomed ROC curve focusing on low π^- fake rates for various energy thresholds	69
5.27 ROC curves for charged tracks for different energy thresholds across different p_T ranges.	72
5.28 Muon's 7×7 pixel images at various beam backgrounds.	74
5.29 Pion's 7×7 pixel images at various beam background levels	74
5.30 ROC curves for different level of beam backgrounds trained with $BG \times 1$	75
5.31 ROC curves for different level of beam backgrounds trained with $BG \times 0$	76
5.32 ROC curves for different level of beam backgrounds trained with $BG \times 2$	77
5.33 ROC curves for different level of beam backgrounds trained with $BG \times 5$	78
5.34 AUC values for the CNN model trained and tested on $BG \times N$ for positive charged tracks.	79

5.35 AUC values for the CNN model trained and tested on $BG \times N$ for negative charged tracks.	79
6.1 Data/MC comparison for total invariant mass	84
6.2 Data/MC comparison for p_T of muons	85
6.3 Data/MC comparison for θ_{ID} of muons	86
6.4 Data/MC comparison for ϕ_{ID} of muons	87
6.5 Comparison of the fitted D^0 mass distributions for MC and data.	90
6.6 Mass distribution of D^0 mesons in the MC simulation of $c\bar{c}$ events, overlaid with data for comparison	91
6.7 Data/MC comparison for p_T of pions	92
6.8 Data/MC comparison for θ_{ID} of pions	93
6.9 Data/MC comparison for ϕ_{ID} of pions	94
6.10 Distribution of CNN outputs for μ^+ and μ^- from the radiative dimuon channel	96
6.11 Distribution of CNN outputs for π^+ and π^- from the D^* channel	97
6.12 ROC curves for various PID methods in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$	99
6.13 ROC curves for various PID methods in three transverse momentum ranges	100
6.14 ROC curves for various energy thresholds in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for positive charged muons and pions	105
6.15 ROC curves for various energy thresholds in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for negative charged muons and pions	106
6.16 ROC curves for various beam backgrounds in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for positive charged muons and pions	108
6.17 ROC curves for various beam backgrounds in the transverse momentum range $0.28 \leq p_T < 1.0 \text{ GeV}/c$ for negative charged muons and pions	109
6.18 Muons' energy depositions in pixel images for MC and data	110
6.19 Distribution of distances of muons to the nearest track	111
6.20 Distribution of distances of muons to the nearest photon	112
6.21 Distribution of CNN outputs for μ^+ and μ^- from the radiative dimuon channel after excluding muon tracks with a nearby photon	113
6.22 ROC curves for charged tracks for all and isolated muons across different p_T ranges.	114

List of Tables

5.1 Energy statistics in all pixels of 7×7 images for events generated under BG \times 1.	45
5.2 Differences between raw and preprocessed MC single track events	48
5.3 Optimal hyperparameter values identified through the optimization process.	51
5.4 μ^+ efficiencies at fixed π^+ fake rates and π^+ fake rates at fixed μ^+ efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for different CNN scenarios.	58
5.5 μ^- efficiencies at fixed π^- fake rates and π^- fake rates at fixed μ^- efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for different CNN scenarios.	59
5.6 μ^+ efficiencies at fixed π^+ fake rates and π^+ fake rates at fixed μ^+ efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for different PID methods.	62
5.7 μ^- efficiencies at fixed π^- fake rates and π^- fake rates at fixed μ^- efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for different PID methods.	63
5.8 Percentage improvements of the CNN over BDT and default PID methods based on AUC values.	65
5.9 Descriptive statistics for AUC values of CNN and BDT methods in different p_T and θ_{ID} ranges.	65
5.10 μ^+ efficiencies at fixed π^+ fake rates and π^+ fake rates at fixed μ^+ efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for various energy thresholds.	70
5.11 μ^- efficiencies at fixed π^- fake rates and π^- fake rates at fixed μ^- efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for various energy thresholds.	71
5.12 AUC values for different energy thresholds across various p_T ranges	71
5.13 μ^+ efficiencies at fixed π^+ fake rates and π^+ fake rates at fixed μ^+ efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/ c for various eam background levels.	76

5.14 μ^- efficiencies at fixed π^- fake rates and π^- fake rates at fixed μ^- efficiencies in the transverse momentum range $0.28 \leq p_T < 1.0$ GeV/c for various eam background levels.	77
6.1 MC signal samples used for $D^{*+} \rightarrow D^0(\rightarrow K^-\pi^+)\pi^+$ and $e^+e^- \rightarrow \mu^+\mu^-\gamma$ channels.	82
6.2 Post-fit values of the model for the D^0 mass distribution	89
6.3 Comparison of μ^\pm and π^\pm counts between data and MC	95
6.4 Comparison of μ efficiencies at fixed π fake rates for $0.28 \leq p_T < 1.0$ GeV/c	101
6.5 Comparison of π fake rates at fixed μ efficiencies for $0.28 \leq p_T < 1.0$ GeV/c	101
6.6 Comparison of μ efficiencies at fixed π fake rates for $0.28 \leq p_T < 0.52$ GeV/c	102
6.7 Comparison of π fake rates at fixed μ efficiencies for $0.28 \leq p_T < 0.52$ GeV/c	102
6.8 Comparison of μ efficiencies at fixed π fake rates for $0.52 \leq p_T < 0.76$ GeV/c	103
6.9 Comparison of π fake rates at fixed μ efficiencies for $0.52 \leq p_T < 0.76$ GeV/c	103
6.10 Comparison of μ efficiencies at fixed π fake rates for $0.76 \leq p_T < 1.0$ GeV/c	104
6.11 Comparison of π fake rates at fixed μ efficiencies for $0.76 \leq p_T < 1.0$ GeV/c	104

Acronyms

SM	Standard Model
ECL	Electromagnetic Calorimeter
PID	Particle Identification
ML	Machine Learning
CNN	Convolutional Neural Network
MC	Monte Carlo
QFT	Quantum Field Theory
QED	Quantum Electrodynamics
QCD	Quantum Chromodynamics
LFU	Lepton Flavour Universality
CKM	Cabibbo–Kobayashi–Maskawa
CM	Center-of-Mass
HVP	Hadronic Vacuum Polarization
HLbL	Hadronic Light-by-Light scattering
MIPs	Minimum Ionizing Particles
HER	High-Energy Ring
LER	Low-Energy Ring
IP	Interaction Point
PXD	Pixel Detector
SVD	Silicon Vertex Detector
VXD	Vertex Detector

CDC	Central Drift Chamber
TOP	Time-Of-Propagation Counter
PMT	Photomultiplier Tube
MCP	Micro-channel Plate
ARICH	Aerogel Ring-Imaging Čherenkov Detector
KLM	K_L^0 and Muon Detector
DAQ	Data Acquisition System
RPC	Resistive-plate Chamber
basf2	Belle II Analysis Software Framework
Geant4	GEometry ANd Tracking version 4
SR	Synchrotron Radiation
CsI(Tl)	Thallium-doped Caesium Iodide
FWD	Forward Endcap
BWD	Backward Endcap
FPGA	Field Programmable Gate Array
BDTs	Boosted Decision Trees
ReLU	Rectified Linear Unit
MSE	Mean Squared Error
SGD	Stochastic Gradient Descent
AdaGrad	Adaptive Gradient Algorithm
RMSProp	Root Mean Square Propagation
NaN	Not a Number
FNN	Feed-forward neural network
TPE	Tree-structured Parzen Estimator
ROC	Receiver Operating Characteristic
TPR	True Positive Rate
FPR	False Positive Rate

AUC	Area Under the Curve
FC	Fully Connected layer
ISR	Initial State Radiation
PSD	Pulse-Shape Discrimination
GNNs	Graph Neural Networks

Bibliography

- [1] E Kou et al. “The Belle II Physics Book”. *Progress of Theoretical and Experimental Physics* 2019.12 (Dec. 2019), p. 123C01. ISSN: 2050-3911. DOI: [10.1093/ptep/ptz106](https://doi.org/10.1093/ptep/ptz106) (cit. on pp. [1](#), [7–9](#), [15–23](#), [26](#)).
- [2] A. Narimani Charan. “Particle identification with the Belle II calorimeter using machine learning”. *Journal of Physics: Conference Series* (2023). DOI: [10.1088/1742-6596/2438/1/012111](https://doi.org/10.1088/1742-6596/2438/1/012111) (cit. on pp. [2](#), [31](#), [137](#)).
- [3] A. Novosel et al. “Identification of light leptons and pions in the electromagnetic calorimeter of Belle II”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1056 (2023), p. 168630. ISSN: 0168-9002. DOI: [10.1016/j.nima.2023.168630](https://doi.org/10.1016/j.nima.2023.168630) (cit. on pp. [2](#), [7](#), [137](#)).
- [4] M. Planck. “Über das Gesetz der Energieverteilung im Normalspectrum”. *Annalen der Physik* 309.3 (1901), pp. 553–563. DOI: [10.1002/andp.19013090310](https://doi.org/10.1002/andp.19013090310) (cit. on p. [3](#)).
- [5] A. Einstein. “Zur Elektrodynamik bewegter Körper”. *Annalen der Physik* 322.10 (1905), pp. 891–921. DOI: [10.1002/andp.19053221004](https://doi.org/10.1002/andp.19053221004) (cit. on p. [3](#)).
- [6] N. Bohr. “I. On the constitution of atoms and molecules”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 26.151 (1913), pp. 1–25. DOI: [10.1080/14786441308634955](https://doi.org/10.1080/14786441308634955) (cit. on p. [3](#)).
- [7] E. Nöther. “Invariante Variationsprobleme”. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257. URL: <http://eudml.org/doc/59024> (cit. on p. [3](#)).
- [8] L. de Broglie. “XXXV. A tentative theory of light quanta”. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 47.278 (1924), pp. 446–458. DOI: [10.1080/14786442408634378](https://doi.org/10.1080/14786442408634378) (cit. on p. [3](#)).
- [9] W. Heisenberg. “Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik”. *Zeitschrift für Physik* 43.3-4 (Mar. 1927), pp. 172–198. DOI: [10.1007/bf01397280](https://doi.org/10.1007/bf01397280) (cit. on p. [3](#)).
- [10] E. Schrödinger. “Quantisierung als Eigenwertproblem”. *Annalen der Physik* 384.4 (1926), pp. 361–376. DOI: [10.1002/andp.19263840404](https://doi.org/10.1002/andp.19263840404) (cit. on p. [3](#)).

- [11] P. A. M. Dirac. “The quantum theory of the electron”. *Proceedings of The Royal Society A: Mathematical, Physical and Engineering Sciences* 117 (1928), pp. 610–624. URL: <https://api.semanticscholar.org/CorpusID:54061357> (cit. on p. [3](#)).
- [12] M. Born. “Zur Quantenmechanik der Stoßvorgänge”. *Zeitschrift für Physik* 37.12 (Dec. 1926), pp. 863–867. ISSN: 0044-3328. DOI: [10.1007/BF01397477](https://doi.org/10.1007/BF01397477) (cit. on p. [3](#)).
- [13] W. Pauli. “Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren”. *Zeitschrift für Physik* 31.1 (Feb. 1925), pp. 765–783. ISSN: 0044-3328. DOI: [10.1007/BF02980631](https://doi.org/10.1007/BF02980631) (cit. on pp. [3](#), [4](#)).
- [14] E. Fermi. “Tentativo di una Teoria Dei Raggi β ”. *Il Nuovo Cimento (1924-1942)* 11.1 (Jan. 1934), pp. 1–19. ISSN: 1827-6121. DOI: [10.1007/BF02959820](https://doi.org/10.1007/BF02959820) (cit. on p. [3](#)).
- [15] R. P. Feynman. “Space-Time Approach to Quantum Electrodynamics”. *Phys. Rev.* 76 (6 Sept. 1949), pp. 769–789. DOI: [10.1103/PhysRev.76.769](https://doi.org/10.1103/PhysRev.76.769) (cit. on p. [3](#)).
- [16] J. Schwinger. “Quantum Electrodynamics. I. A Covariant Formulation”. *Phys. Rev.* 74 (10 Nov. 1948), pp. 1439–1461. DOI: [10.1103/PhysRev.74.1439](https://doi.org/10.1103/PhysRev.74.1439) (cit. on p. [3](#)).
- [17] S. Tomonaga. “On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields”. *Progress of Theoretical Physics* 1.2 (Aug. 1946), pp. 27–42. ISSN: 0033-068X. DOI: [10.1143/PTP.1.27](https://doi.org/10.1143/PTP.1.27) (cit. on p. [3](#)).
- [18] M. Gell-Mann. “A schematic model of baryons and mesons”. *Physics Letters* 8.3 (1964), pp. 214–215. ISSN: 0031-9163. DOI: [10.1016/S0031-9163\(64\)92001-3](https://doi.org/10.1016/S0031-9163(64)92001-3) (cit. on p. [4](#)).
- [19] S. Weinberg. “A Model of Leptons”. *Physical Review Letters* 19.21 (1967), pp. 1264–1266. DOI: [10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264) (cit. on pp. [4](#), [6](#)).
- [20] A. Salam. “Weak and Electromagnetic Interactions”. *Conf. on High Energy Physics, N. N. Bogoliubov et al., eds., JINR, Dubna, USSR* (1968), pp. 367–377 (cit. on pp. [4](#), [6](#)).
- [21] S. L. Glashow. “Partial-symmetries of weak interactions”. *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: [10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2) (cit. on pp. [4](#), [6](#)).
- [22] M. Thomson. *Modern particle physics*. Cambridge University Press, 2013. ISBN: 978-1-107-03426-6. DOI: [10.1017/CB09781139525367](https://doi.org/10.1017/CB09781139525367) (cit. on pp. [4](#), [5](#), [8](#), [12](#), [13](#)).
- [23] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. *Physics Letters B* 716 (2012), pp. 1–29. DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020) (cit. on p. [5](#)).
- [24] S. Chatrchyan et al. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. *Physics Letters B* 716 (2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021) (cit. on p. [5](#)).
- [25] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321) (cit. on p. [5](#)).

- [26] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. *Physical Review Letters* 13.16 (1964), pp. 508–509. DOI: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508) (cit. on p. [5](#)).
- [27] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: [10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585) (cit. on p. [5](#)).
- [28] *Standard Model*. Accessed: July 3, 2023. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (cit. on p. [5](#)).
- [29] D. Griffiths. *Introduction to Elementary Particles*. Wiley-VCH, 2008 (cit. on pp. [6](#), [7](#)).
- [30] R. Feynman. *Quantum Electrodynamics*. Addison-Wesley, 1962 (cit. on p. [6](#)).
- [31] F. Halzen and A. D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. John Wiley & Sons, 1984 (cit. on p. [7](#)).
- [32] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: [10.1103/PhysRevLett.13.138](https://doi.org/10.1103/PhysRevLett.13.138) (cit. on p. [7](#)).
- [33] M. Kobayashi and T. Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. *Progress of Theoretical Physics* 49 (1973), pp. 652–657. DOI: [10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652) (cit. on p. [7](#)).
- [34] A. J. Bevan et al. “The Physics of the B Factories”. *The European Physical Journal C* 74.11 (Nov. 2014), p. 3026. ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-014-3026-9](https://doi.org/10.1140/epjc/s10052-014-3026-9) (cit. on pp. [7](#), [8](#)).
- [35] B. R. Martin and G. Shaw. *Particle Physics*. John Wiley & Sons, 2017 (cit. on pp. [8](#), [10](#)).
- [36] A. Aloisio et al. “The hadronic cross section measurement at KLOE”. *Nuclear Physics B - Proceedings Supplements* 144 (2005). TAU 2004, pp. 231–237. ISSN: 0920-5632. DOI: [10.1016/j.nuclphysbps.2005.02.032](https://doi.org/10.1016/j.nuclphysbps.2005.02.032) (cit. on p. [8](#)).
- [37] F. Ambrosino et al. “Measurement of $\sigma(e^+e^- \rightarrow \pi^+\pi^-\gamma(\gamma))$ and the dipion contribution to the muon anomaly with the KLOE detector”. *Physics Letters B* 670.4 (2009), pp. 285–291. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2008.10.060](https://doi.org/10.1016/j.physletb.2008.10.060) (cit. on p. [8](#)).
- [38] The KLOE-2 collaboration et al. “Combination of KLOE $\sigma(e^+e^- \rightarrow \pi^+\pi^-\gamma(\gamma))$ measurements and determination of $a_\mu^{\pi^+\pi^-}$ in the energy range $0.10 < s < 0.95$ GeV²”. *Journal of High Energy Physics* 2018.3 (Mar. 2018), p. 173. ISSN: 1029-8479. DOI: [10.1007/JHEP03\(2018\)173](https://doi.org/10.1007/JHEP03(2018)173) (cit. on p. [8](#)).
- [39] B. et al. Aubert. “Precise Measurement of the $e^+e^- \rightarrow \pi^+\pi^-(\gamma)$ Cross Section with the Initial State Radiation Method at BABAR”. *Physical Review Letters* 103.23 (2009). DOI: [10.1103/PhysRevLett.103.231801](https://doi.org/10.1103/PhysRevLett.103.231801) (cit. on p. [8](#)).
- [40] J. P. et al. Lees. “Precise measurement of the $e^+e^- \rightarrow \pi^+\pi^-(\gamma)$ cross section with the initial-state radiation method at BABAR”. *Phys. Rev. D* 86 (3 Aug. 2012), p. 032013. DOI: [10.1103/PhysRevD.86.032013](https://doi.org/10.1103/PhysRevD.86.032013) (cit. on p. [8](#)).

- [41] M. Ablikim et al. “Measurement of the $e^+e^- \rightarrow \pi^+\pi^-$ cross section between 600 and 900 MeV using initial state radiation”. *Physics Letters B* 753 (2016), pp. 629–638. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2015.11.043](https://doi.org/10.1016/j.physletb.2015.11.043) (cit. on p. [8](#)).
- [42] T. Xiao et al. “Precision measurement of the hadronic contribution to the muon anomalous magnetic moment”. *Phys. Rev. D* 97 (3 Feb. 2018), p. 032012. DOI: [10.1103/PhysRevD.97.032012](https://doi.org/10.1103/PhysRevD.97.032012) (cit. on p. [8](#)).
- [43] R.R Akhmetshin et al. “Update: A reanalysis of hadronic cross section measurements at CMD-2”. *Physics Letters B* 578.3 (2004), pp. 285–289. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2003.10.108](https://doi.org/10.1016/j.physletb.2003.10.108) (cit. on p. [8](#)).
- [44] M. N. Achasov et al. “Update of the $e^+e^- \rightarrow \pi^+\pi^-$ cross section measured by the spherical neutral detector in the energy region $400 < \sqrt{s} < 1000$ MeV”. *Journal of Experimental and Theoretical Physics* 103.3 (Sept. 2006), pp. 380–384. ISSN: 1090-6509. DOI: [10.1134/S106377610609007X](https://doi.org/10.1134/S106377610609007X) (cit. on p. [8](#)).
- [45] Friedrich Jegerlehner. *The Anomalous Magnetic Moment of the Muon*. Springer International Publishing, 2017. DOI: [10.1007/978-3-319-63577-4](https://doi.org/10.1007/978-3-319-63577-4) (cit. on p. [9](#)).
- [46] C. Grupen and B. Schwartz. *Particle Detectors*. Cambridge University Press, 2008 (cit. on pp. [9](#)–[13](#)).
- [47] H. Bethe. “Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie”. *Annalen der Physik* 397.3 (1930), pp. 325–400. DOI: [10.1002/andp.19303970303](https://doi.org/10.1002/andp.19303970303) (cit. on p. [9](#)).
- [48] H. Bethe. “Bremsformel für Elektronen relativistischer Geschwindigkeit”. *Zeitschrift für Physik* 76.5–6 (May 1932), pp. 293–299. ISSN: 1434-601X. DOI: [10.1007/bf01342532](https://doi.org/10.1007/bf01342532) (cit. on p. [9](#)).
- [49] F. Bloch. “Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie”. *Annalen der Physik* 408.3 (1933), pp. 285–320. DOI: [10.1002/andp.19334080303](https://doi.org/10.1002/andp.19334080303) (cit. on p. [9](#)).
- [50] C. et al. Patrignani. “Review of Particle Physics”. *Chin. Phys. C* 40.10 (2016), p. 100001. DOI: [10.1088/1674-1137/40/10/100001](https://doi.org/10.1088/1674-1137/40/10/100001) (cit. on pp. [9](#)–[12](#)).
- [51] W. R. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach*. Springer, 1987 (cit. on pp. [11](#), [12](#)).
- [52] R. Wigmans. *Calorimetry: Energy Measurement in Particle Physics*. Oxford University Press, 2000 (cit. on p. [11](#)).
- [53] C. W. Fabjan and F. Gianotti. “Calorimetry for particle physics”. *Reviews of Modern Physics* 75.4 (Oct. 2003), pp. 1243–1286. DOI: [10.1103/RevModPhys.75.1243](https://doi.org/10.1103/RevModPhys.75.1243) (cit. on p. [11](#)).
- [54] K. Akai, K. Furukawa, and H. Koiso. “SuperKEKB collider”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 907 (2018), pp. 188–199. ISSN: 0168-9002. DOI: [10.1016/j.nima.2018.08.017](https://doi.org/10.1016/j.nima.2018.08.017) (cit. on pp. [15](#), [16](#)).

- [55] T. Abe et al. “Belle II Technical Design Report”. (Nov. 2010). arXiv: [1011.0352](https://arxiv.org/abs/1011.0352) [[physics.ins-det](https://arxiv.org/abs/1011.0352)] (cit. on pp. [15](#), [17](#), [19](#), [20](#), [22](#), [23](#)).
- [56] E. Hemsing et al. “Beam by design: Laser manipulation of electrons in modern accelerators”. *Rev. Mod. Phys.* 86 (3 July 2014), pp. 897–941. DOI: [10.1103/RevModPhys.86.897](https://doi.org/10.1103/RevModPhys.86.897) (cit. on p. [17](#)).
- [57] *Belle II archive*. Accessed: October 1, 2023. 2023. URL: <https://www.belle2.org/archives/> (cit. on p. [18](#)).
- [58] T. Kuhr et al. “The Belle II Core Software”. *Computing and Software for Big Science* 3.1 (2018). ISSN: 2510-2044. DOI: [10.1007/s41781-018-0017-9](https://doi.org/10.1007/s41781-018-0017-9) (cit. on p. [21](#)).
- [59] S. Agostinelli et al. “Geant4 - a simulation toolkit”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8) (cit. on p. [21](#)).
- [60] J. Allison et al. “Geant4 developments and applications”. *IEEE Transactions on Nuclear Science* 53.1 (2006), pp. 270–278. DOI: [10.1109/TNS.2006.869826](https://doi.org/10.1109/TNS.2006.869826) (cit. on p. [21](#)).
- [61] H. Ikeda. “Development of the CsI(Tl) calorimeter for the measurement of CP violation at KEK *B*-factory”. PhD thesis. Nara Women’s University, Japan, 1999 (cit. on pp. [23](#)–[26](#)).
- [62] M. De Nuccio. “Search for Axion-Like Particles produced in e^+e^- collisions and photon energy resolution studies at Belle II”. PhD thesis. Universität Hamburg, 2021. URL: <https://ediss.sub.uni-hamburg.de/handle/ediss/9323> (cit. on pp. [24](#), [25](#)).
- [63] M. Milesi, J. Tan, and P. Urquijo. “Lepton identification in Belle II using observables from the electromagnetic calorimeter and precision trackers”. *EPJ Web of Conferences* 245 (2020), p. 06023. ISSN: 2100-014X. DOI: [10.1051/epjconf/202024506023](https://doi.org/10.1051/epjconf/202024506023) (cit. on pp. [26](#)–[30](#)).
- [64] V. Lakshminarayanan and A. Fleck. “Zernike polynomials: a guide”. *Journal of Modern Optics* 58.7 (Apr. 2011), pp. 545–561. DOI: [10.1080/09500340.2011.554896](https://doi.org/10.1080/09500340.2011.554896) (cit. on p. [29](#)).
- [65] D.N. Brown, J. Ilic, and G.B. Mohanty. “Extracting longitudinal shower development information from crystal calorimetry plus tracking”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 592.3 (2008), pp. 254–260. ISSN: 0168-9002. DOI: [10.1016/j.nima.2008.04.029](https://doi.org/10.1016/j.nima.2008.04.029) (cit. on p. [29](#)).
- [66] Andreas Hocker et al. “TMVA - Toolkit for Multivariate Data Analysis”. (Mar. 2007). arXiv: [physics/0703039](https://arxiv.org/abs/physics/0703039) (cit. on p. [29](#)).
- [67] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, 2006 (cit. on pp. [31](#), [33](#), [35](#), [37](#), [39](#)).
- [68] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. URL: <http://www.deeplearningbook.org> (cit. on pp. [31](#)–[35](#), [37](#)–[39](#)).

- [69] Y. LeCun et al. “Gradient-based learning applied to document recognition”. *Proceedings of the IEEE* 86 (1998), pp. 2278–2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791) (cit. on p. [31](#)).
- [70] A. Reynolds. *Convolutional Neural Networks (CNNs)*. 2019. URL: <https://anhreynolds.com/notes/convolutional-neural-networks.html> (visited on 06/25/2023) (cit. on p. [32](#)).
- [71] MathWorks. *Introduction to Deep Learning: What Are Convolutional Neural Networks?* Accessed: 2023-06-25. URL: <https://de.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html> (cit. on p. [34](#)).
- [72] F. Chollet. *Deep Learning with Python*. Manning Publications Company, 2018 (cit. on pp. [36](#), [37](#)).
- [73] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning representations by back-propagating errors”. *Nature* 323.6088 (Oct. 1986), pp. 533–536. ISSN: 1476-4687. DOI: [10.1038/323533a0](https://doi.org/10.1038/323533a0) (cit. on pp. [36](#), [37](#)).
- [74] M. A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. URL: <http://neuralnetworksanddeeplearning.com> (cit. on pp. [37](#), [38](#)).
- [75] L. Bottou. “Large-Scale Machine Learning with Stochastic Gradient Descent”. *Proceedings of COMPSTAT 2010* (2010), pp. 177–186. DOI: [10.1007/978-3-7908-2604-3_16](https://doi.org/10.1007/978-3-7908-2604-3_16) (cit. on p. [38](#)).
- [76] D. P. Kingma and J. Ba. “Adam: A Method for Stochastic Optimization”. *CoRR* abs/1412.6980 (2014). URL: <https://api.semanticscholar.org/CorpusID:6628106> (cit. on p. [38](#)).
- [77] V. Dumoulin and F. Visin. “A guide to convolution arithmetic for deep learning”. *ArXiv* abs/1603.07285 (2016). URL: <https://api.semanticscholar.org/CorpusID:6662846> (cit. on p. [39](#)).
- [78] J. Bergstra and Y. Bengio. “Random Search for Hyper-Parameter Optimization”. *Journal of Machine Learning Research* 13.10 (2012), pp. 281–305. URL: <http://jmlr.org/papers/v13/bergstra12a.html> (cit. on p. [39](#)).
- [79] N. Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: <http://jmlr.org/papers/v15/srivastava14a.html> (cit. on p. [39](#)).
- [80] L. Prechelt. “Early Stopping-But When?”. *Neural Networks: Tricks of the Trade*. 1996 (cit. on p. [39](#)).
- [81] W. Rawat and Z. Wang. “Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review”. *Neural Computation* 29.9 (Sept. 2017), pp. 2352–2449. DOI: [10.1162/neco_a_00990](https://doi.org/10.1162/neco_a_00990) (cit. on p. [43](#)).
- [82] C. Shorten and T. M. Khoshgoftaar. “A survey on Image Data Augmentation for Deep Learning”. *Journal of Big Data* 6.1 (July 2019). DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0) (cit. on p. [43](#)).

- [83] Y. Bengio, A. Courville, and P. Vincent. “Representation Learning: A Review and New Perspectives”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.8 (2013), pp. 1798–1828. DOI: [10.1109/TPAMI.2013.50](https://doi.org/10.1109/TPAMI.2013.50) (cit. on p. [43](#)).
- [84] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2019 (cit. on p. [49](#)).
- [85] Adam Paszke et al. “PyTorch: An Imperative Style, High-Performance Deep Learning Library”. *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf> (cit. on p. [51](#)).
- [86] Pang-Ning Tan. “Receiver Operating Characteristic”. *Encyclopedia of Database Systems*. Boston, MA: Springer US, 2009, pp. 2349–2352. ISBN: 978-0-387-39940-9. DOI: [10.1007/978-0-387-39940-9_569](https://doi.org/10.1007/978-0-387-39940-9_569) (cit. on p. [55](#)).
- [87] S. Jadach, B. F. L. Ward, and Z. Waś. “The Precision Monte Carlo event generator KK for two fermion final states in e^+e^- collisions”. *Comput. Phys. Commun.* 130 (2000), pp. 260–325. DOI: [10.1016/S0010-4655\(00\)00048-5](https://doi.org/10.1016/S0010-4655(00)00048-5) (cit. on p. [82](#)).
- [88] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024) (cit. on p. [82](#)).
- [89] D. J. Lange. “The EvtGen particle decay simulation package”. *Nucl. Instrum. Methods Phys. Res. A* 462.1 (2001), pp. 152–155. DOI: [10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4) (cit. on p. [82](#)).
- [90] A. Martini. *Description of the muonID algorithm and performances study on $e^+e^- \rightarrow \mu^+\mu^-\gamma$ channel*. 2020 (cit. on p. [83](#)).
- [91] P.A. et al. Zyla. “Review of Particle Physics”. *PTEP* 2020.8 (2020), p. 083C01. DOI: [10.1093/ptep/ptaa104](https://doi.org/10.1093/ptep/ptaa104) (cit. on p. [88](#)).
- [92] R. L. et al. Workman. “Review of Particle Physics”. *PTEP* 2022 (2022), p. 083C01. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097) (cit. on p. [89](#)).
- [93] S. Longo and J. M. Roney. “Hadronic vs. electromagnetic pulse shape discrimination in CsI(Tl) for high energy physics experiments”. *Journal of Instrumentation* 13.03 (Mar. 2018), P03018–P03018. DOI: [10.1088/1748-0221/13/03/p03018](https://doi.org/10.1088/1748-0221/13/03/p03018) (cit. on p. [116](#)).
- [94] S. Longo et al. “CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 982 (2020), p. 164562. ISSN: 0168-9002. DOI: [10.1016/j.nima.2020.164562](https://doi.org/10.1016/j.nima.2020.164562) (cit. on p. [116](#)).

Disclaimer

The CNN PID methodology, as presented in this thesis, was proposed to me by my supervisor, Prof. Dr. Torben Ferber. The analytical framework itself was formulated by me, and the implementation of the CNN PID method within the Belle II software framework was performed by me.

For the purpose of training the CNN models, which is detailed in Chapter 5, I took on the task of generating simulated data. This was accomplished by leveraging the capabilities of the Belle II software suite. The beam background overlays were produced by the Belle II collaboration. The Belle II collaboration, of which I am part since 2019, designed, built, and runs the Belle II experiment.

Moving forward to the validation phase of the CNN PID method with data, as explained in Chapter 6, the dataset employed was provided by the Belle II collaboration. The electron and positron beams are provided by the SuperKEKB collider. The data used for the CNN validation were collected in 2020 and 2021.

The culmination of this research journey resulted in the publication of an article, authored by myself, which briefly presents the findings and methodology of the CNN PID method. This article was published in the Journal of Physics: Conference Series in February 2023 [2]. Additionally, I have contributed as an author to a supplementary article that extends upon our work, published in the Nuclear Instruments and Methods in Physics Research Section A in November 2023 [3]. It is worth noting that my participation in various other publications, as part of the Belle II collaboration, extends beyond the scope of this thesis [4].

¹For an list of publications, please consult <https://inspirehep.net/literature?sort=mostrecent&size=25&page=1&q=A%20Narimani%20Charan>.

Acknowledgements

I wish to express my gratitude to Prof. Dr. Torben Ferber and Dr. Carsten Niebuhr for affording me the invaluable opportunity to embark on my PhD journey within their research group. Torben, your mentorship has imparted to me not only profound physics knowledge but also essential insights into effective project management and strategies. The lessons learned during this endeavor have far exceeded my initial expectations.

My deepest appreciation goes to my wife, Fateme. Words alone cannot adequately convey my gratitude for your support.

I extend my sincere thanks to my parents, whose love and encouragement have played an instrumental role in shaping the person I am today.

I am indebted to Dr. Alberto Martini, Dr. Michel Hernandez Villanueva, Dr. Filippo Dattola, and Marcel Hohmann for their meticulous proofreading of my thesis.

A multitude of individuals have contributed to my success on this journey. Foremost among them are my colleagues and friends: Dr. Alberto Martini, Dr. Michel Hernandez Villanueva, Dr. Michael De Nuccio, Sascha Dreyer, Dr. Filippo Dattola, Dr. Cyrille Praz, Marcel Hohmann, Dr. Hendrikas Svidras, Anja Novosel, Dr. Marco Milesi, Merle Graf-Schreiber, and Anselm Baur. Your selfless assistance and patience in helping me through this transformative period in my life have been a beacon of inspiration.

Finally, I extend my gratitude to other colleagues and teammates, Dr. Sam Cunliffe and Dr. Savino Longo, for their invaluable counsel and support during this endeavor.

Declaration

I hereby declare upon oath that I have not submitted a dissertation with the same research topic to an academic higher education institution that had been already accepted or evaluated as insufficient in an earlier doctoral procedure.

Ich versichere, dass ich keine Dissertation mit dem gleichen Forschungsthema schon einmal in einem fruheren Promotionsverfahren an einer wissenschaftlichen Hochschule eingereicht habe, die angenommen oder als ungenugend beurteilt worden ist.

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

I hereby declare upon oath that the dissertation submitted in electronic form and the printed bound copy of the dissertation submitted for archiving are identical.

Ich versichere an Eides statt, dass die in elektronischer Form eingereichte Dissertation und das zur Archivierung eingereichte gedruckte gebundene Exemplar der Dissertation identisch sind.



Abtin Narimani Charan
Hamburg, 14 February 2024

