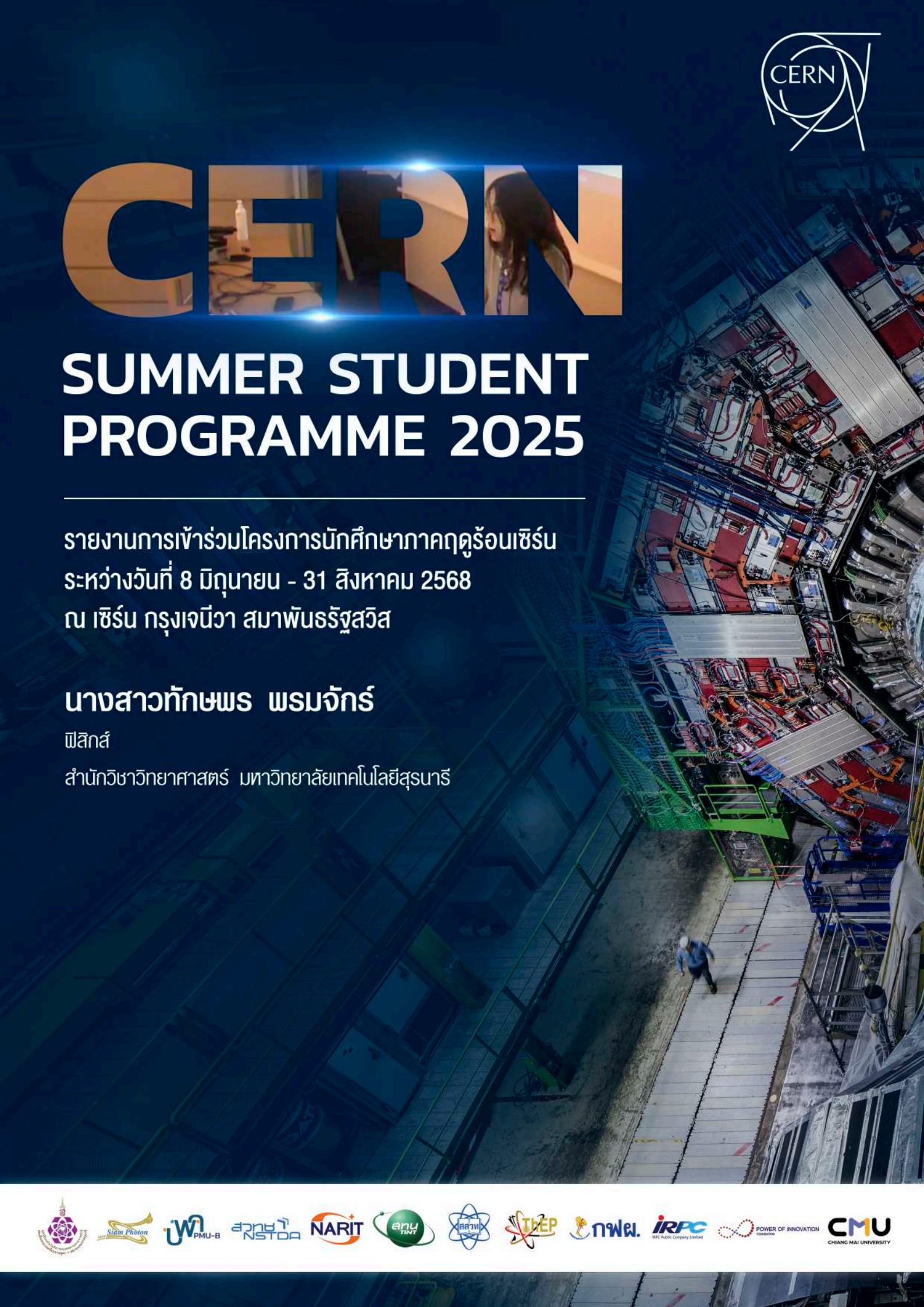


CERN

SUMMER STUDENT PROGRAMME 2025

รายงานการเข้าร่วมโครงการนักศึกษาภาคฤดูร้อนเซิร์น
ระหว่างวันที่ 8 มิถุนายน - 31 สิงหาคม 2568
ณ เซิร์น กรุงเจนีวา สมาพันธรัฐสวิส

นางสาวทักษพร พรมจักร์

ฟิสิกส์

สำนักวิชาวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีสุรนารี

Preface

Since my high school years, I have been deeply interested in high energy physics, which led me to actively participate in numerous summer schools and academic programs dedicated to this field. These early experiences not only strengthened my passion for particle physics but also provided me with the foundation and motivation to pursue research in this area throughout my university life.

One of my long-standing academic goals has been to take part in the CERN Summer Student Program, an opportunity regarded worldwide as a unique platform for young researchers to learn, contribute, and grow in the heart of modern physics. Having now achieved this goal, I feel both honored and inspired to share the knowledge and experiences I gained during my time at CERN.

This report, which includes both my project report and a diary of my daily experiences, is written with the hope of offering valuable insights to future students. Just as I was inspired by reading the reports of previous participants, I aspire that my reflections may likewise encourage and guide others who wish to join this remarkable program. Beyond serving as a record of my own academic journey, I hope this report highlights the collaborative spirit, intellectual challenges, and profound excitement that come with being part of the international scientific community at CERN.

Taksaporn Promjak

Acknowledgement

This participation would not have been possible without the support of the Thai-CERN Collaboration Program under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn, The Information Technology Foundation under the Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn, and the Synchrotron Light Research Institute (Public Organization), who kindly guided and assisted me throughout the entire process of the program. I am also deeply grateful to the Program Management Unit for Human Resources & Institutional Development,Research and Innovation (PMU-B) for financially supporting my participation. I would like to sincerely thank the distinguished selection committee for giving me this valuable opportunity to learn and gain such an extraordinary experience.

I would like to extend my special thanks to Dr. Norraphat Srimanobhas for organizing and assisting with a project collaboration with CERN. I am also very grateful to Prof. Dr. Yupeng Yan and Dr. Chinorat Kobdaj for encouraging me to apply to the program and for their valuable guidance and suggestions. The wonderful journey at CERN could neither have begun nor been completed without their support.

I would also like to acknowledge Jordy Butter, Chen Chen, Matthew William Kenzie, Yingrui Hou, and Laetitia Marie Guerry for their mentorship, which not only deepened my understanding of physics but also sharpened my research and problem-solving skills in many aspects.

Finally, I would like to thank my family and friends for their unwavering encouragement and support in both my studies and my life.

CERN Summer Student Report
29 August 2025
taksaporn.promjak@cern.ch

Search for the Unobserved $B_s^0 \to K^- \rho^+$ Decay and π^0 Reconstruction Performance Investigation at LHCb During Run 3

Author: Taksaporn Promjak
Supervisors: Jordy Butter, Chen Chen
Matthew William Kenzie, Yingrui Hou, Laetitia Marie Guerry
CERN, CH-1211 Geneva, Switzerland

Keywords: π^0 **Reconstruction, Electromagnetic calorimeter (ECAL), boosted decision trees (BDT)**

Abstract

This work presents a study of neutral pion π^0 reconstruction in the context of charmless hadronic B_s^0 meson decays at LHCb, with a particular focus on the search for the yet-unobserved decay $B_s^0 \to K^-\pi^+\pi^0$. This channel, expected to proceed predominantly through the intermediate resonance $B_s^0 \to K^-\rho^+$, provides a sensitive probe to flavor dynamics, CP violation, and possible contributions beyond the Standard Model. The analysis exploits both Run 2 (2018) and Run 3 (2024), specifically block7 and block8 from 2024, datasets, comparing detector performance before and after the LHCb upgrade. In Run 2, the π^0 invariant mass peak is well reconstructed and consistent with the PDG value, enabling stable reconstruction of B^0 and B_s^0 candidates. In Run 3, however, broader and slightly shifted π^0 and B^0 mass distributions are observed, indicating the challenges introduced by higher occupancies, software-based triggering, and possible mis-calibration of the upgraded electromagnetic calorimeter (ECAL). A multivariate approach based on boosted decision trees (XGBoost) is employed to enhance signal purity, with the event selection process ensuring robust training and systematic control. The results emphasize the central role of π^0 reconstruction quality in B-meson analyses and highlight the necessity of improved photon energy calibration and region-dependent corrections in Run3. Looking ahead, refined π^0 calibration in Run 3, together with improved energy reconstruction and PID corrections, will be essential for a precise determination of the $B_s^0 \to K^-\pi^+\pi^0$ branching fraction.

Project report

1 Introduction

1.1 Motivation

The study of charmless hadronic B-meson decays plays an important role in testing the Standard Model description of flavor dynamics and in constraining possible contributions from physics beyond the Standard Model. In this context, the decay $B_s^0 \to K^-\pi^+\pi^0$, which is expected to proceed predominantly through the intermediate resonance $B_s^0 \to K^-\rho^+$, is of particular interest. This channel has not yet been observed, and its measurement would provide valuable input for theoretical frameworks based on flavor symmetries and QCD factorization, as well as for amplitude analyses and studies of direct CP violation in the B_s^0 system [1]. The related decay $B^0 \to K^+\rho^-$, which has the same final-state topology and is Cabibbo-favored relative to the B_s^0 mode with a well-measured branching fraction', serves as an important control channel to validate the analysis strategy.

A particular feature of this channel is the presence of a neutral pion in the final state. The π^0 decays almost exclusively to two photons and, owing to its neutral charge, cannot be directly tracked in the detector. Its reconstruction relies entirely on the electromagnetic calorimeter (ECAL) [2], making the performance of photon and π^0 reconstruction in terms of efficiency, resolution, and energy calibration a critical factor for the sensitivity of the analysis. The higher occupancies, the aging of the LHCb ECAL, and the transition to a fully software-based trigger and alignment system in Run 3 add further challenges for photon-rich final states, underlining the importance of a careful characterization of π^0 performance.

This work is therefore motivated both by the physics interest in a previously unobserved B_s^0 decay and by the methodological requirement to improve the reconstruction of neutral pions in LHCb Run 3 conditions.

1.2 Physics context and goals

The primary objective of this work is the search for the unobserved decay $B_s^0 \to K^-\pi^+\pi^0$, which is expected to proceed predominantly through the intermediate resonance $B_s^0 \to K^-\rho^+$. Should evidence for this decay be established, its branching fraction will be measured and compared to theoretical expectations. To support this search, the related decay $B^0 \to K^+\rho^-$, which shares the same final-state topology and has a significantly larger branching fraction, is employed as a control channel for validating the analysis strategy and assessing systematic effects. In addition, an analysis of the channel $B \to hh\pi^0$ is being carried out using Run 2 (2018) data, following the same selection procedure with preselection requirements, veto cuts, and validation steps. This parallel study provides an important benchmark for testing the neutral pion reconstruction and the overall strategy in a well-established dataset before applying it to the B_s^0 search.

A central goal of the analysis is to evaluate the reconstruction performance of neutral pions in LHCb Run 3 conditions. Since π^0 mesons are reconstructed exclusively from photons detected in the ECAL, their resolution and efficiency play a critical role in determining the overall sensitivity of the measurement. This study therefore aims not only to optimize the selection and background suppression for the signal mode but also to provide a quantitative assessment of π^0 reconstruction performance, which has broader implications for analyses involving neutral final states at LHCb.

1.3 Structure of this report

This report is organized as follows. Section 2 provides the theoretical and experimental background relevant to neutral pion reconstruction, including π^0properties, ECAL performance, and the main reconstruction challenges. Section 3 describes the datasets used in this analysis, together with the simulation samples and truth information. Section 4 outlines the event selection strategy, including preselection, veto cuts, and sideband definitions for validation. Section 5 presents the machine learning approach, covering the choice of classifier, input variables, and overtraining checks. Section 6 discusses the reconstruction studies and results, focusing first on π^0 reconstruction and then on B-meson mass reconstruction under different selection conditions. Finally, Section 7 summarizes the conclusions of this work and provides an outlook for future studies.

2 Theoretical and Experimental Background

2.1 Neutral pion properties and $\pi^0 \to \gamma\gamma$

The neutral pion (π^0) is the lightest meson, with quantum numbers $I^G(J^{PC}) = 1^-(0^{-+})$, and a mass of 134.9768 ± 0.0005 MeV/c^2. It has a lifetime of $(8.43 \pm 0.13) \times 10^{-17}$ s and decays almost exclusively $(98.833 \pm 0.034)\%$ into two photons via the electromagnetic interaction [3]. Because the π^0 carries no electric charge, it leaves no direct track in the detector. Its reconstruction is therefore fully dependent on detecting its decay photons with the ECAL.

In the context of this analysis, these properties are particularly relevant because the performance of π^0 reconstruction depends strongly on the calorimeter's photon energy and position resolution, as well as on its ability to separate nearby showers. The small mass of the π^0 means that its decay photons are often highly collimated at high momentum, leading to overlapping or merged clusters in the ECAL. This complicates the identification of π^0 candidates and directly impacts the mass resolution and background rejection in searches for $B_s^0 \to K^-\pi^+\pi^0$. Consequently, understanding the decay kinematics and detector response to $\pi^0 \to \gamma\gamma$ is essential for developing an efficient multivariate classifier and evaluating the systematic uncertainties associated with neutral pion reconstruction.

2.2 Resolved vs. merged photons and ECAL

The neutral pion decays promptly to two photons, $\pi^0 \to \gamma\gamma$. Depending on the pion momentum, the photon pair can appear in two distinct topologies:

Resolved π^0
Resolved π^0s occur when the two photons from the π^0 decay are sufficiently separated at the ECAL to form two distinct clusters. This typically happens when the pion's momentum is low to moderate. The two photons strike the ECAL at different locations, creating two separate energy deposits.

A resolved photon pair is a cleaner signal because it allows for direct reconstruction of the parent pion's mass and kinematics. The detector's ability to resolve these photons depends on its granularity—the smaller the individual ECAL cells (Shashlik-like, consist of scintillator layer and Pb layer), the better its spatial and energy resolution.

Merged π^0

Merged π^0s occur when the π^0 is produced with very high momentum. In this case, the two photons travel at such a small opening angle to each other that they hit the ECAL very close together. The detector registers them as a single, large, and sometimes oddly shaped cluster. It's impossible to distinguish the two photons individually.

This topology is more challenging to analyze because the standard two-photon reconstruction is not possible. Special algorithms are needed to identify these single, merged clusters and interpret them as a high-momentum pion. The ability to detect merged photons is crucial for analyzing high-energy processes where parent particles are often produced with significant boost.

LHCb ECAL

The LHCb ECAL is designed to measure the energy and position of electrons and photons through the detection of electromagnetic showers. It is segmented into three regions: the inner, middle, and outer regions. This design reflects the variation in particle flux across the detector acceptance, with the finest granularity in the inner region near the beam pipe, and progressively coarser segmentation in the middle and outer regions.

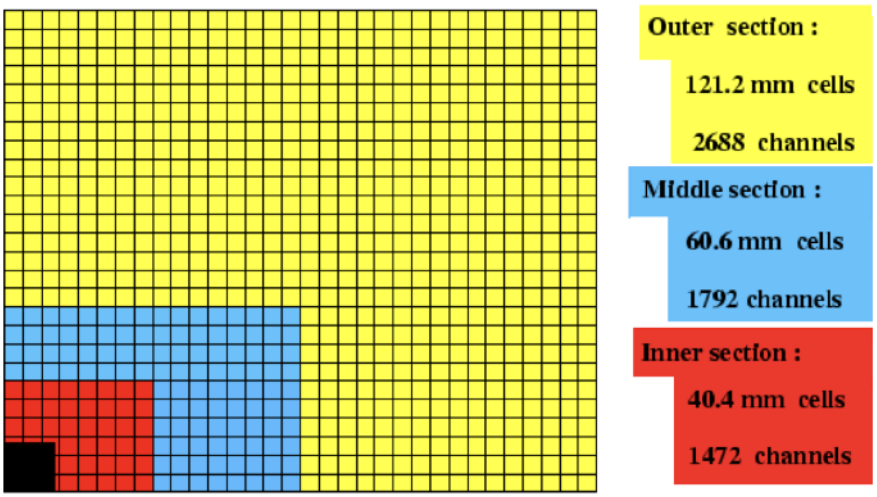


Figure 1: Calorimeter cells segmentation of the ECAL [4]

For photon reconstruction, this segmentation is particularly relevant for the decay $\pi^0 \to \gamma\gamma$. Depending on their separation, the two decay photons may be reconstructed as distinct clusters (resolved π^0) or as a single merged cluster. At high momentum, the photons become increasingly collimated, making merged reconstruction more likely in regions of coarser granularity, while the inner region provides improved separation power. This distinction between resolved and merged topologies affects both the reconstruction efficiency and the invariant mass resolution of π^0 candidates, and is therefore explicitly considered in the analysis.

2.3 Reconstruction challenges relevant to this analysis

In the analysis of the decay $B_s^0 \to K^-\rho^+ (\rho^+ \to \pi^+\pi^0)$, several reconstruction challenges arise that directly affect the sensitivity and precision of the measurement. A central difficulty lies in the reconstruction of neutral pions, which rely exclusively on photon detection in the ECAL and are thus subject to limitations in energy resolution, shower overlap, and photon conversions. These effects propagate to the invariant mass resolution of both the π^0 and the B_s^0 candidate, complicating the separation of signal from background. Furthermore, efficient particle identification of charged tracks, suppression of combinatorial and overlapping backgrounds, and

the performance of the Run 3 software trigger introduce additional sources of complexity. The following subsections provide a detailed discussion of these challenges and their implications for the present analysis.

Neutral pion (π^0) reconstruction

The decay $B_s^0 \to K^-\rho^+$ involves the intermediate state $\rho^+ \to \pi^+\pi^0$, with the neutral pion promptly decaying into two photons. Unlike charged particles, π^0 mesons cannot be directly tracked and must be reconstructed entirely from their photon decay products in the ECAL. This dependence on calorimeter performance introduces challenges due to limited energy and position resolution, photon shower overlap, and inefficiencies from photon conversions. Accurate π^0 reconstruction is essential for fully reconstructing the decay chain, improving the B_s mass resolution, and enhancing signal-to-background separation.

Photon reconstruction and ECAL limitations

Since π^0 reconstruction depends solely on photon detection, the performance of the ECAL plays a critical role. Photons may convert before reaching the calorimeter, degrading energy measurement and complicating the reconstruction. Moreover, the calorimeter resolution itself limits the precision with which photon energies and angles can be measured, worsening the invariant mass resolution of π^0 candidates. These effects lead to high background rates from combinatorial photon pairs, which can mimic genuine π^0 signals.

Charged particle tracking and particle identification

While the neutral pion is the most challenging particle in the final state, accurate reconstruction of the charged particles (K^- and π^+) is equally important. The tracking detectors must provide good momentum resolution despite multiple scattering and detector inefficiencies. Particle identification (PID), built from the combined information of multiple subdetectors (RICH, tracking dE/dx, calorimeters, and the muon system), is required to cleanly separate kaons from pions, since even small misidentification rates can lead to significant background in rare decay searches like $B_s^0 \to K^-\rho^+$.

Combinatorial background and overlapping channels

The presence of neutral pions greatly increases the presence of combinatorial background contributions, as many random photon pairs can form fake π^0 candidates. Advanced selection techniques, such as boosted decision trees (BDT), are therefore essential to improve background rejection while maintaining signal efficiency. In addition, overlapping channels such as $B^0 \to K^+\rho^-$ or $B^0 \to K^+\pi^-\pi^0$, which look like B_s^0 have higher yields, can obscure the B_s signal. In particular, because the final state contains a π^0, the B-candidate mass resolution is degraded, and the long tail of the B^0 mass distribution can overlap with and obscure the B_s^0 peak.

Mass resolution and kinematic fitting

The relatively poor resolution of π^0 reconstruction propagates into the reconstructed B_s candidate mass, broadening the peak and making it harder to distinguish signal from background. Kinematic fitting techniques, which constrain intermediate resonances and apply vertex constraints, can partially mitigate this issue [5]. However, their effectiveness is still limited by the inherent uncertainties in photon energy measurement and π^0 reconstruction.

3 Data and Simulation Samples

3.1 Run 2 and Run 3 datasets

The analysis uses LHCb data from two periods: (i) Run 2 2018 pp collisions at $\sqrt{s} = 13$ TeV, and (ii) Run 3 2024 pp collisions at $\sqrt{s} = 13.6$ TeV, specifically Blocks 7 and 8. The 2018 sample provides provides a well-understood reference with stable detector performance and hardware-based triggering, serving as a benchmark for validating reconstruction methods. The 2024 sample corresponds to the upgraded detector with a fully software-based trigger and higher occupancies, introducing additional challenges for photon and π^0 reconstruction. The comparison between the two runs enables assessment of reconstruction performance before and after the detector upgrade.In this study we focus on the resolved π^0 category events where the two photons form distinct ECAL clusters, while merged π^0 candidates are not considered in the baseline results.

3.2 Truth information (ID vs TRUEID)

In simulation samples, we have access not only to the reconstructed particle **ID** (assigned by the reconstruction software) but also to the generator-level identity, **TRUEID**, and its parents. Truth matching makes use of this extra information to ensure that the candidates used in training and validation correspond exactly to the desired decay chain.

3.3 Event Selection

Truth matching and BKGCAT

Truth matching requires one hadron to be reconstructed as a charged kaon and the other hadron to be reconstructed as an oppositely charged pion. and the π^0 is the mother particle of two photons. ID criteria are applied to select the case that both photons originate from the same true π^0, and that this π^0 is the daughter of the correct B hadron. Especially in MC simulation (Monte Carlo simulation), TRUEID is used to check that the hadrons match the kaon/pion hypothesis and that both photons are children of a true π^0 whose mother is the expected B.

In addition, the **BKGCAT** (Background Category) variable provides in the LHCb simulation framework is used to categorize candidates according to their origin, distinguishing between true signal decays, combinatorial background, and cross-feed from other B-hadron decays. This allows a cleaner definition of training samples and a more precise estimation of background contributions.

This strict definition of "ground truth" ensures that our multivariate classifier is trained on true signal decays, so that efficiency measurements and systematic studies are not biased by mis-reconstructed or mis-identified particles. Without truth matching, the classifier might inadvertently learn to separate "clean" from "dirty" candidates inside MC, which does not reflect the real background composition in data.

Truth matching (Run 2 data)	Truth matching (Run 3 data)
h1_ID = 321	|hp_ID| = 321
h2_ID = −211	|hm_ID| = 211
Pi0_ID = 111	pi0_ID = 111
Pi0_gam1_ID = 22	g1_ID = 22
Pi0_gam2_ID = 22	g2_ID = 22

Table 1: Truth matching and BKGCAT applied in Run 2 data(left) and Run 3 data(right).

Truth matching (Run 2 MC)	Truth matching (Run 3 MC)
h1_ID = 321	|hp_ID| = 321
h2_ID = −211	|hm_ID| = 211
Pi0_ID = 111	pi0_ID = 111
Pi0_gam1_ID = 22	g1_ID = 22
Pi0_gam2_ID = 22	g2_ID = 22
h1_TRUEID = 321	|hp_TRUEID| = 321
h2_TRUEID = −211	|hm_TRUEID| = 211
B_BKGCAT = 0 | (B_BKGCAT = 50 & g1_TRUEID = 22 & g2_TRUEID = 22 & pi0_TRUEID = 111 & (pi0_MC_MOTHER_ID = -511 | pi0_MC_MOTHER_ID = 511))	B_BKGCAT = 0 | (B_BKGCAT = 50 & g1_TRUEID = 22 & g2_TRUEID = 22 & pi0_TRUEID = 111 & (pi0_MC_MOTHER_ID = -511 | pi0_MC_MOTHER_ID = 511))

Table 2: Truth matching and BKGCAT applied in Run 2 MC(left) and Run 3 MC(right).

While truth matching is essential for building clean MC samples, it cannot be applied to real data. To bridge this, we use preselection criteria: a physics-motivated set of requirements applied uniformly to both data and simulation.

Preselection cuts

The online selection ensures that the B candidate forms a good-quality displaced vertex pointing back to the primary interaction, that its decay products are well-identified within the detector subsystems, and that their kinematics lie within the acceptance of the LHCb detector.

The preselection defines the baseline quality filter applied before any machine learning or final mass fits. The goal is to reject candidates that are clearly mis-reconstructed or physically implausible, while retaining as much genuine signal as possible. These requirements are motivated by the detector geometry, the physics of weak decays, and the known performance of tracking, calorimetry, and particle-identification algorithms.

In summary, the online selection together with the preselection filters enforce realistic physics topology and detector quality, balance the signal-to-background ratio, and ensure that the variables used in multivariate training are well-behaved. They also reduce computational

load by discarding implausible candidates at an early stage, preventing pathological events from biasing classifiers or inflating statistical uncertainties. The same preselection strategy is applied consistently in both Run 2 and Run 3, with only minor technical differences reflecting the updated reconstruction framework.

Preselection (Run 2)
$\text{Prob}(\texttt{hB\_VTXCHI2DOF}) > 0.03$
$\texttt{hB\_ptasy\_2.00} > -0.25$
$\texttt{Pi0\_CL} > 0.20$
$\texttt{Tot\_IsNotE} > 0.10$
$\texttt{IsMuon} \neq 1$
$\log(\texttt{hB\_VTXISODCHI2ONETRACK}) > -0.5$
$\texttt{hB\_MIPCHI2\_PV} < 9$
$\texttt{hasRich} \neq 0$
$2590\ \text{MeV}/c^2 \leq \texttt{h\_P} \leq 100000\ \text{MeV}/c^2$

Preselection (Run 3)
$\text{Prob}(\texttt{B\_VTXCHI2DOF}) > 0.03$
$\texttt{B\_DELTAR\_004\_CC\_CONE\_ASYM\_PT} > -0.25$
$\texttt{Pi0\_CL} > 0.20$
$0 < \texttt{B\_MIN\_BPVIPCHI2} < 40000$
$\texttt{ISMuon} \neq 1$
$\texttt{OWNPVIPCHI2} > 0$
$\texttt{B\_OWNPVIPCHI2} < 10$
$\texttt{PPHASRICH} \neq 0$
$2590\ \text{MeV}/c^2 \leq \texttt{h\_P} \leq 100000\ \text{MeV}/c^2$
$\texttt{hp\_PROBNN\_K} > 0.1$
$\texttt{hm\_PROBNN\_PI} > 0.1$

Table 3: Baseline preselection used in Run 2 (left) and Run 3 (right).

While the preselection strategy removes most misreconstructed candidates, the surviving sample is largely signal-like. Consequently, the limiting backgrounds are not random combinatorial combinations but specific hadronic decays that share the same visible final state. We therefore introduce a set of targeted vetoes to reject these well-identified peaking backgrounds rather than random candidates. For these, we employ veto cuts. **Veto cuts**
Even after preselection, significant background remains from real hadronic decays that share the same final-state topology as the signal. These "peaking backgrounds" are particularly problematic because they form good vertices, have valid PID responses, and occupy similar kinematic regions to the desired B meson decay. As such, they cannot be suppressed by quality requirements alone.

Veto cuts are designed to target these specific reflections by removing candidates whose two-body subsystems are consistent with known resonances, charm decays (D^0, $D^\pm$, $D_s^\pm$) and misidentified two-body B decays, that can mimic the signal structure. By excluding regions around the well-measured masses of these particles, the analysis applies targeted vetoes to remove specific backgrounds that could mimic the signal. This trade-off is essential: without vetoes, charm contributions and other resonant decays would dominate the selected sample, obscuring the rare signal under study.

In short, veto cuts act as a precision filter on top of preselection, ensuring that the surviving candidates cannot be trivially explained by high-yield background channels. They are therefore a critical step in stabilising the analysis and preparing a dataset where multivariate techniques can focus on separating true signal from combinatorial noise, rather than being confounded by resonant look-alikes.

Veto cuts
$\mid m_{K^{\pm}\pi^{\mp}} - m_{D^0} \mid > 30$ MeV/c^2
$\mid m_{K^{\pm}\pi^{\mp}} - m_{B^0} \mid > 50$ MeV/c^2
$\mid m_{\pi^{\pm}\pi^0} - m_{D^{\pm}} \mid > 60$ MeV/c^2
$\mid m_{K^{\pm}\pi^0} - m_{D_s^{\pm}} \mid > 60$ MeV/c^2
$\mid m_{K^{\pm}\pi^0} - m_{B^{\pm}} \mid > 102$ MeV/c^2

Table 4: Veto cuts applied in both Run 2 and Run 3 dataset.

With truth-matched, physics-motivated preselection, and targeted vetoes, the remaining sample is clean enough to apply machine learning. The multivariate classifier then learns to distinguish signal from the dominant combinatorial background.

3.4 Sideband definitions for validation

After applying preselection and veto cuts, the dataset still contains a large amount of combinatorial background. To model this component and provide a background sample for multivariate training, we define mass sidebands around the B-candidate invariant mass. Events in these sidebands are assumed to be kinematically similar to the signal region but lie outside the physical B-mass peak, making them a reliable proxy for combinatorial background. By using sidebands, we avoid relying on simulated samples for combinatorial background, which are not trustworthy, and instead ensure that the BDT learns from data-driven examples of the dominant background sources.

- Run 2 sideband : $hB\_pi0constPVconst\_M > 5600$ MeV
- Run 3 sideband : $B\_OwnPV\_MASS > 5600$ MeV

With the sidebands establishing a data-driven background sample and the truth-matched MC defining the signal, we proceed to train a boosted decision tree (BDT) to further separate signal from the remaining background.

4 Machine Learning Approach

4.1 Classifier choice (BDT/XGBoost) and configuration

The suppression of combinatorial background in the search for $B_s^0 \to K^-\rho^+(\rho^+ \to \pi^+\pi^0)$ requires a multivariate approach, since the discriminating information is distributed across many observables with only moderate individual separation power. A boosted decision tree (BDT), implemented with XGBoost, is employed due to its ability to exploit non-linear correlations, robustness to heterogeneous inputs, and proven performance in high-energy physics analyses [6].

The classifier is trained using simulated signal events and background candidates from mass sidebands, with additional contributions from simulated cross-feed and partially reconstructed decays to improve stability. Input variables are restricted to those with minimal correlation to the reconstructed mass in order to avoid biasing the signal extraction. Hyperparameters such as maximum depth, learning rate, and subsampling are tuned through cross-validation, while overtraining is controlled by comparing training and validation responses. The optimal working point is chosen based on the expected sensitivity of the final mass fit rather than classifier metrics alone.

Since the effectiveness of the classifier depends on the choice of discriminating features, the next subsection introduces the set of input variables employed in the training and explains their physical motivation.

4.2 Input variables

The classifier is trained on a set of variables with low correlation to the reconstructed mass and complementary separation power.

Run 2:

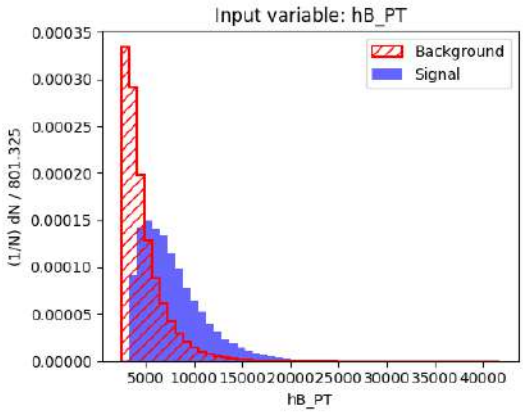


Figure 2: Normalized distribution of the input variable `hB_PT`

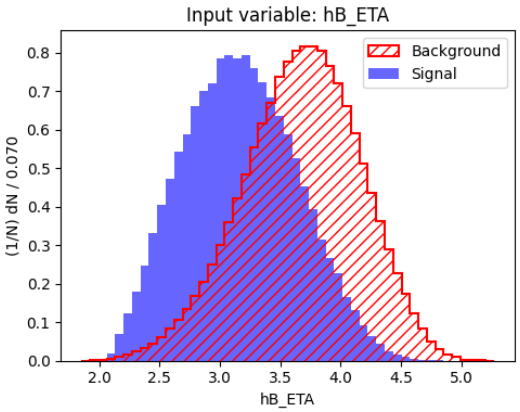


Figure 3: Normalized distribution of the input variable `hB_ETA`

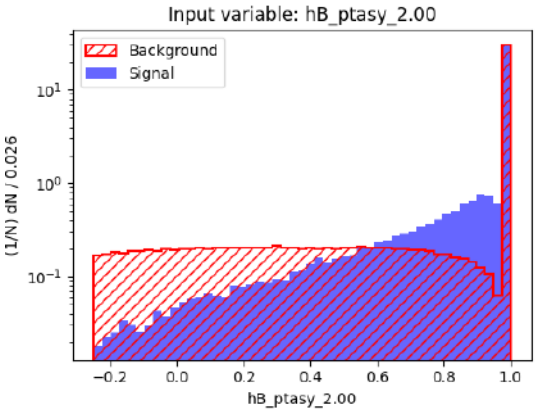


Figure 4: Normalized distribution of the input variable `hB_ptasy_2.00`

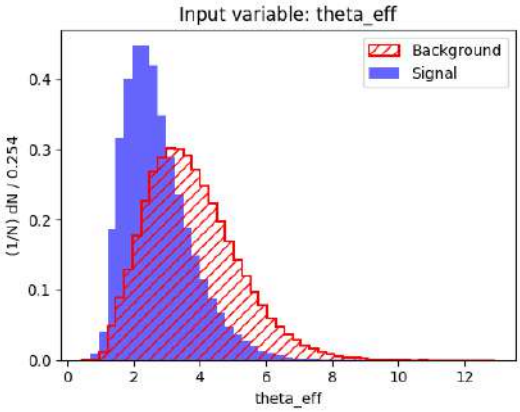


Figure 5: Normalized distribution of the input variable `theta_eff`

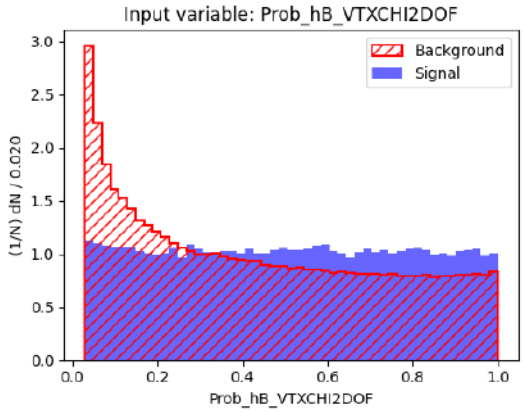


Figure 6: Normalized distribution of the input variable `Prob(hB_VTXCHI2DOF)`

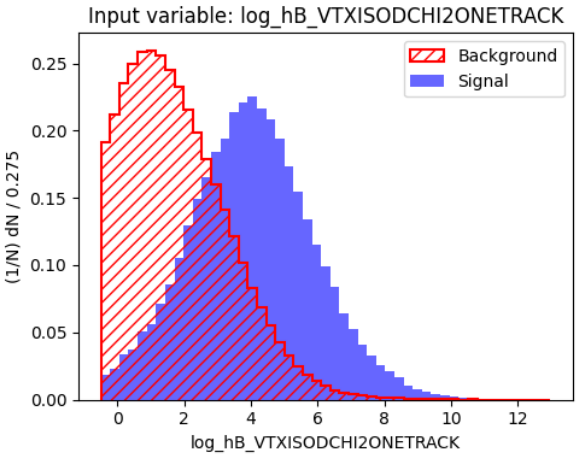


Figure 7: Normalized distribution of the input variable `log(hB_VTXISODCHI2ONETRACK)`

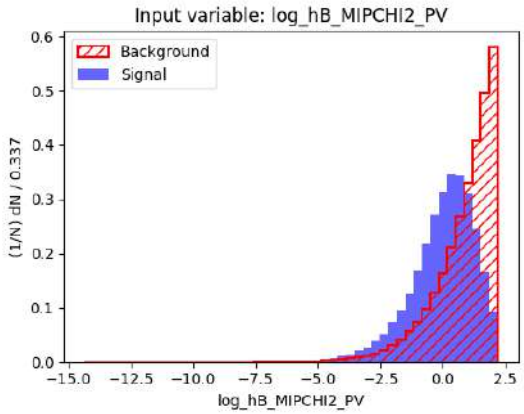


Figure 8: Normalized distribution of the input variable `log(hB_MIPCHI2_PV)`

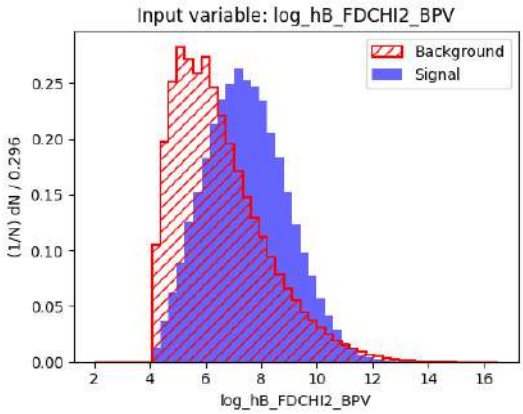


Figure 9: Normalized distribution of the input variable `log(hB_FDCHI2_BPV)`

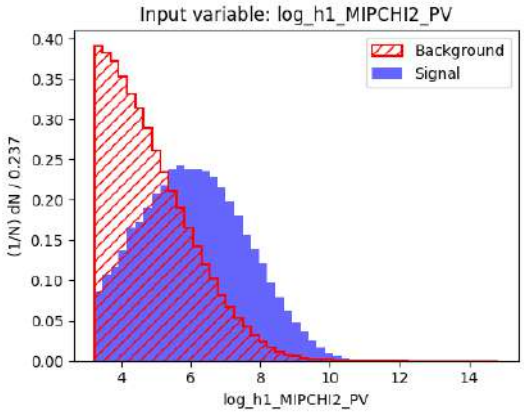


Figure 10: Normalized distribution of the input variable `log(h1_MIPCHI2_PV)`

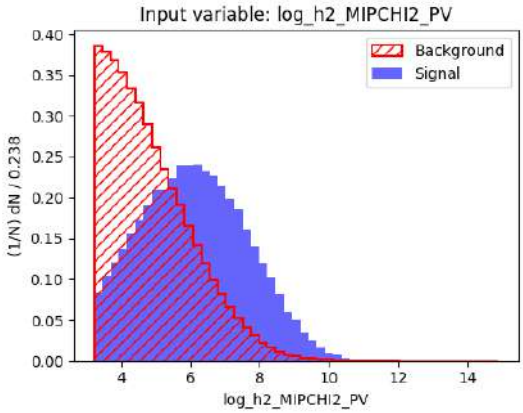


Figure 11: Normalized distribution of the input variable `log(h2_MIPCHI2_PV)`

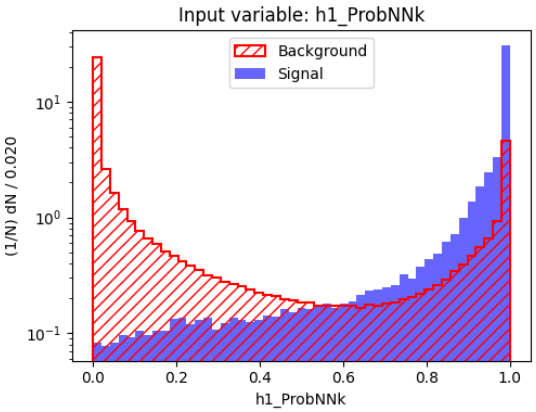


Figure 12: Normalized distribution of the input variable `h1_ProbNNk`

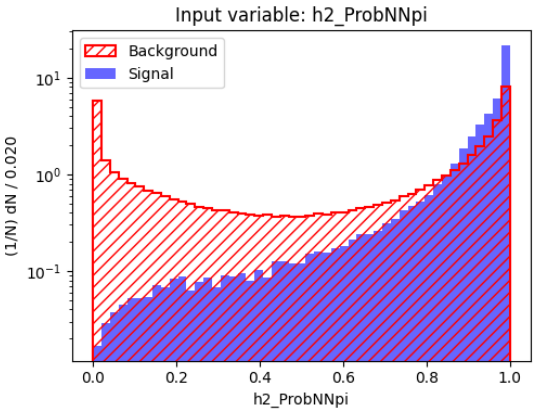


Figure 13: Normalized distribution of the input variable `h2_ProbNNpi`

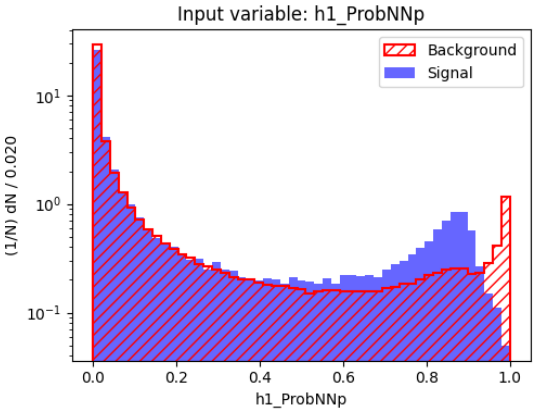


Figure 14: Normalized distribution of the input variable `h1_ProbNNp`

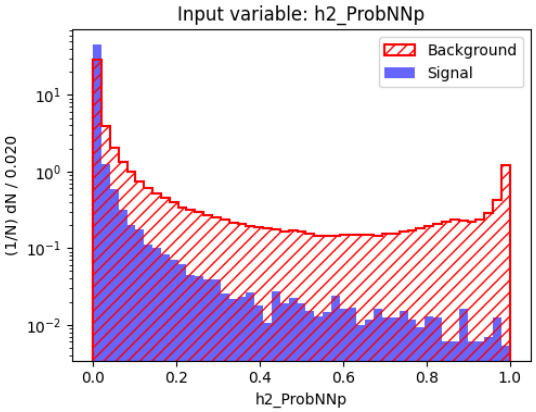


Figure 15: Normalized distribution of the input variable `h2_ProbNNp`

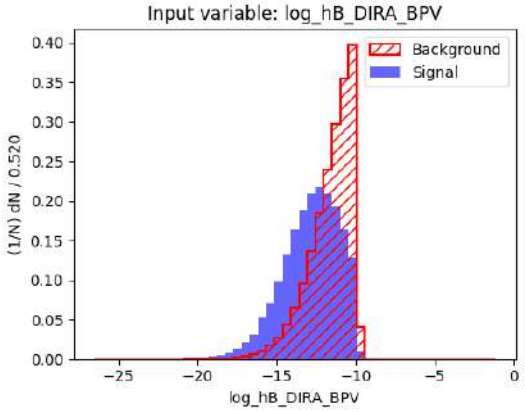


Figure 16: Normalized distribution of the input variable `log(hB_DIRA_BPV)`

- **hB_PT**: Transverse momentum of the B candidate.
- **hB_ETA**: Pseudorapidity (η) of the B candidate flight direction.
- **hB_ptasy_2.00**: Transverse momentum asymmetry of the B candidate within a cone of radius 2.0.
- **theta_eff**: Effective opening angle between the B decay products. The cell size is considered for this variable. (the mattematical expression is shown in appendix B.)
- **Prob(hB_VTXCHI2DOF)**: Vertex fit quality of the B candidate.
- **log(hB_VTXISODCHI2ONETRACK)**: Isolation of the B vertex with respect to nearby tracks.
- **log(hB_MIPCHI2_PV)**: Minimum impact parameter χ^2 of the B candidate with respect to the primary vertex.
- **log(hB_FDCHI2_BPV)**: Flight distance significance of the B candidate from the primary vertex.
- **log(h1_MIPCHI2_PV)**: Impact parameter χ^2 of the first daughter track with respect to the primary vertex.
- **log(h2_MIPCHI2_PV)**: Impact parameter χ^2 of the second daughter track with respect to the primary vertex.
- **h1_ProbNNk**: Probability that the first daughter is identified as a kaon.
- **h2_ProbNNpi**: Probability that the second daughter is identified as a pion.
- **h1_ProbNNp**: Probability that the first daughter is identified as a proton.
- **h2_ProbNNp**: Probability that the second daughter is identified as a proton.
- **log(hB_DIRA_BPV)**: Cosine of the pointing angle between the B candidate momentum and the vector from the primary vertex, indicating alignment of the decay direction.

Run 3:

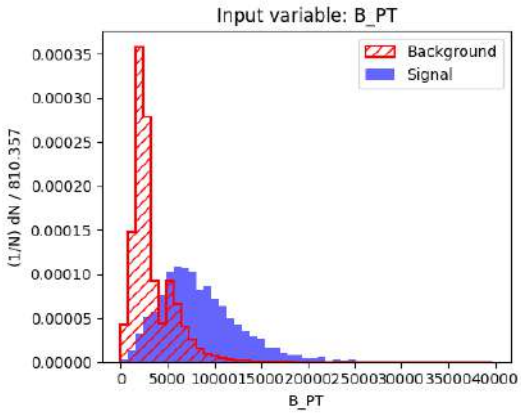


Figure 17: Normalized distribution of the input variable `B_PT`

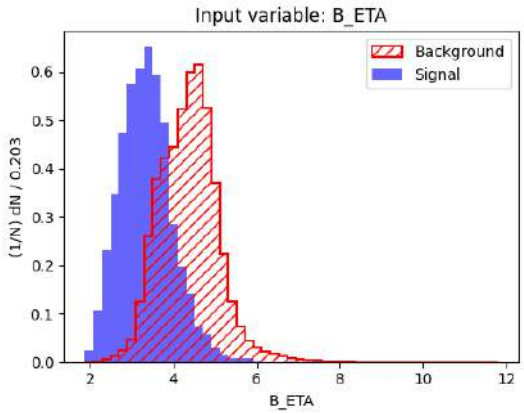


Figure 18: Normalized distribution of the input variable `B_ETA`

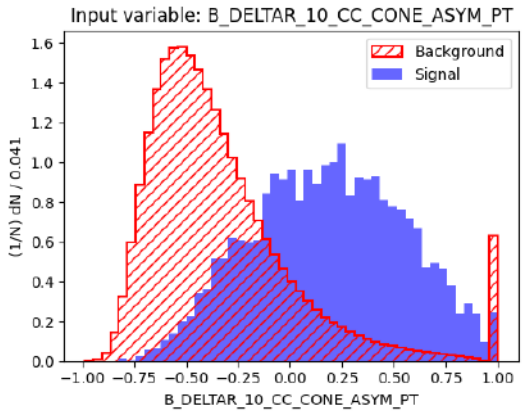


Figure 19: Normalized distribution of the input variable `B_DELTAR_10_CC_CONE_ASYM_PT`

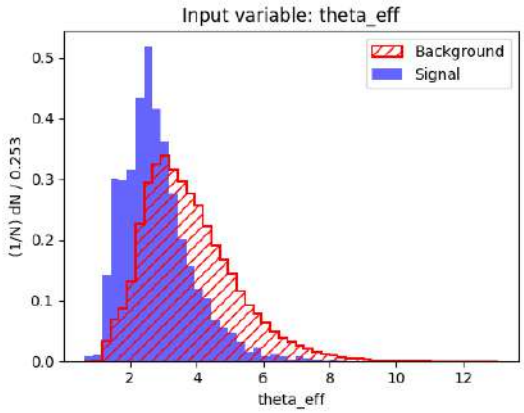


Figure 20: Normalized distribution of the input variable `theta_eff`

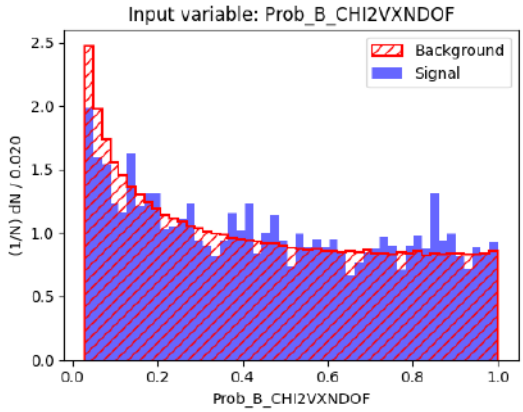


Figure 21: Normalized distribution of the input variable `Prob(B_CHI2VXNDOF)`

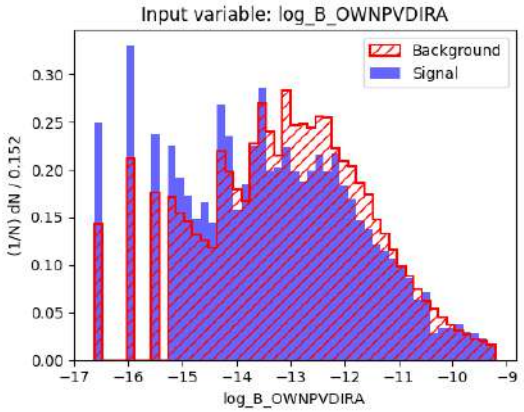


Figure 22: Normalized distribution of the input variable `log(B_OWNPVDIRA)`

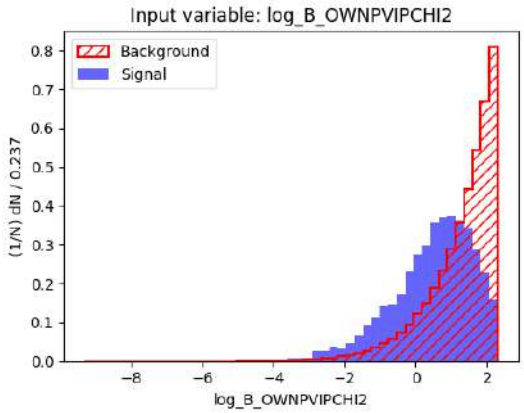


Figure 23: Normalized distribution of the input variable `log(B_OWNPVIPCHI2)`

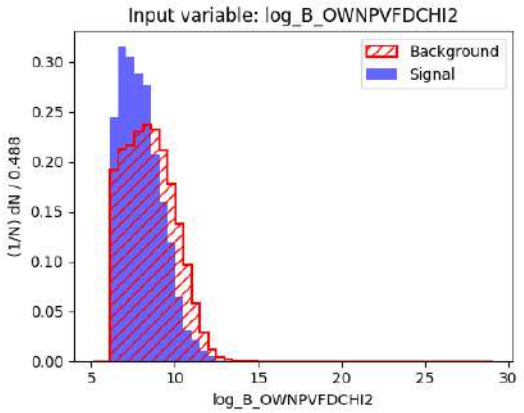


Figure 24: Normalized distribution of the input variable `log(B_OWNPVFDCHI2)`

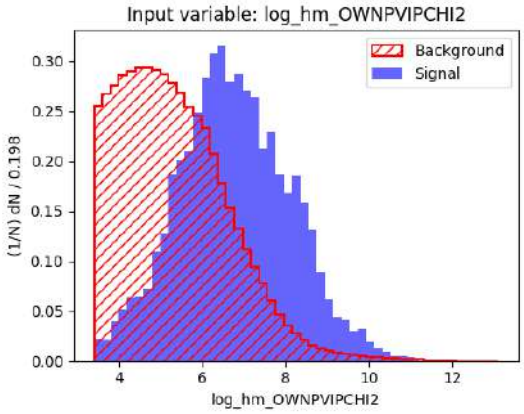


Figure 25: Normalized distribution of the input variable `log(hm_OWNPVIPCHI2)`

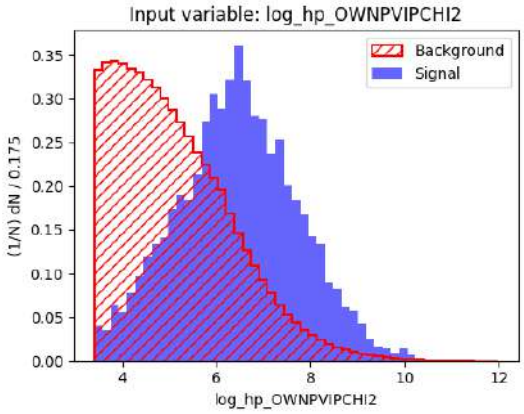


Figure 26: Normalized distribution of the input variable `log(hp_OWNPVIPCHI2)`

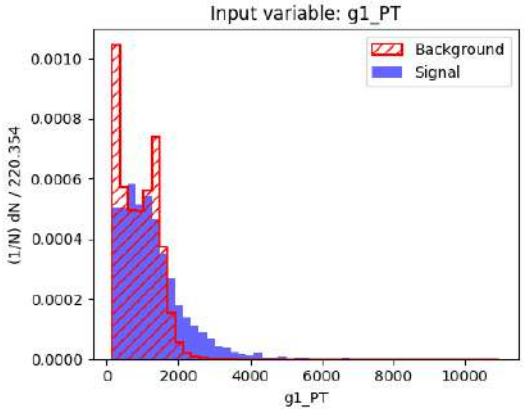


Figure 27: Normalized distribution of the input variable `g1_PT`

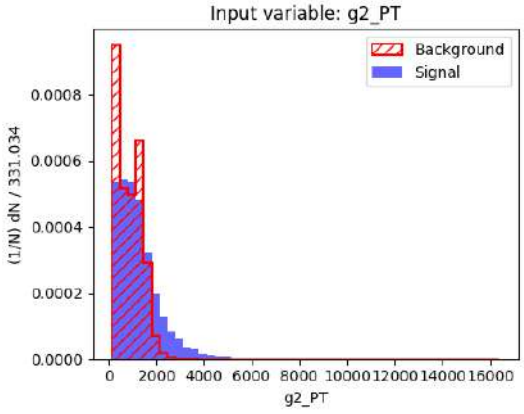


Figure 28: Normalized distribution of the input variable `g2_PT`

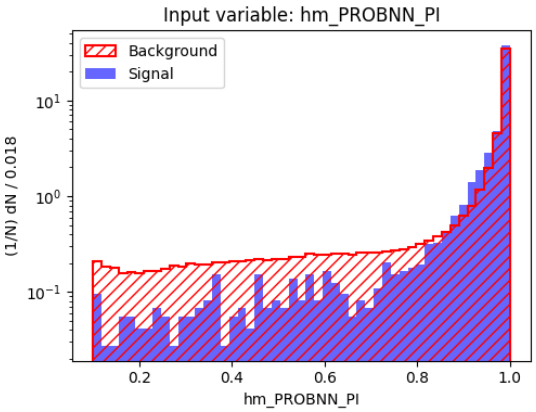


Figure 29: Normalized distribution of the input variable `hm_PROBNN_PI`

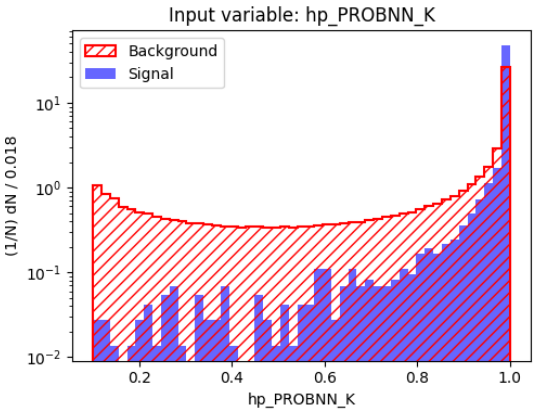


Figure 30: Normalized distribution of the input variable `hp_PROBNN_K`

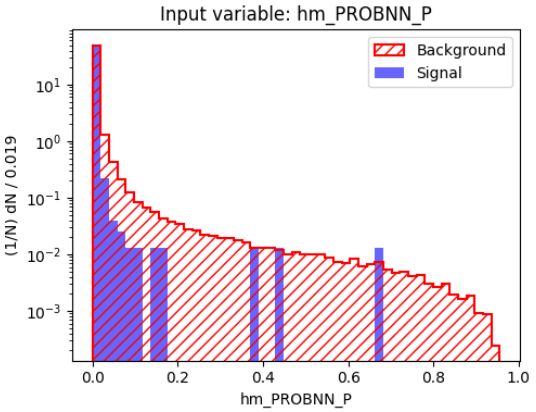


Figure 31: Normalized distribution of the input variable `hm_PROBNN_P`

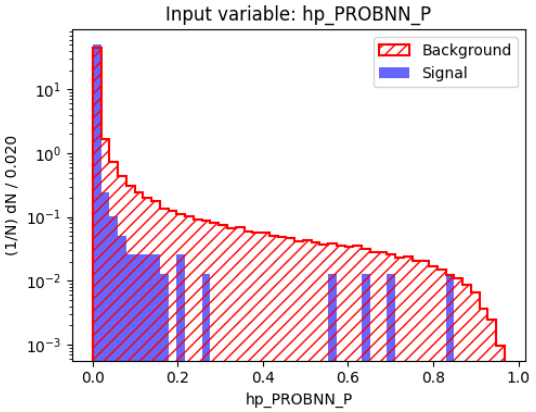


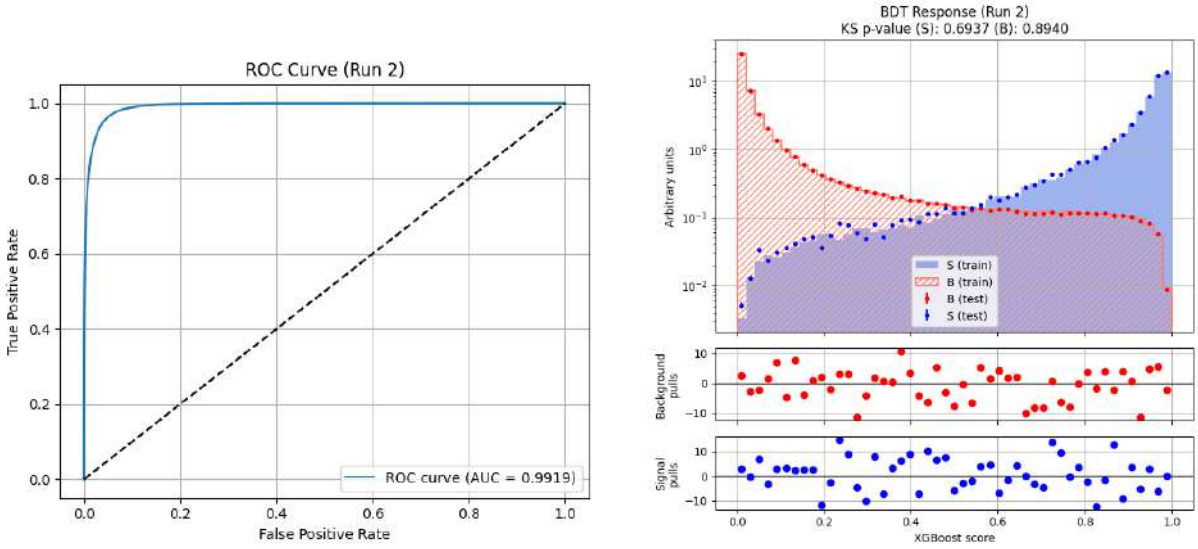
Figure 32: Normalized distribution of the input variable `hp_PROBNN_P`

- **B_PT**: Transverse momentum of the B candidate.
- **B_ETA**: Pseudorapidity (η) of the B candidate.
- **B_DELTAR_10_CC_CONE_ASYM_PT**: Transverse momentum asymmetry of the B within a cone of radius 1.0.
- **theta_eff**: Effective opening angle between the B decay products.
- **Prob(B_CHI2VXNDOF):** Vertex fit quality of the B candidate.
- **log(B_OWNPVDIRA):** Cosine of the pointing angle between the B candidate momentum and the direction from the primary vertex.
- **log(B_OWNPVIPCHI2):** Impact parameter χ^2 of the B candidate with respect to the primary vertex.
- **log(B_OWNPVFDCHI2):** Flight distance significance of the B candidate from the primary vertex.
- **log(hm_OWNPVIPCHI2):** Impact parameter χ^2 of the negative charged daughter track with respect to the primary vertex.

- **log(hp_OWNPVIPCHI2):** Impact parameter χ^2 of the positive charged daughter track with respect to the primary vertex.
- **g1_PT:** Transverse momentum of the first photon (γ_1).
- **g2_PT:** Transverse momentum of the second photon (γ_2).
- **hm_PROBNN_PI:** Probability that the negative charged daughter track is identified as a pion.
- **hp_PROBNN_K:** Probability that the positive charged daughter track is identified as a kaon.
- **hm_PROBNN_P:** Probability that the negative charged daughter track is identified as a proton.
- **hp_PROBNN_P:** Probability that the positive charged daughter track is identified as a proton.

4.3 Overtraining checks: ROC, KS tests, pull distributions and the feature importances

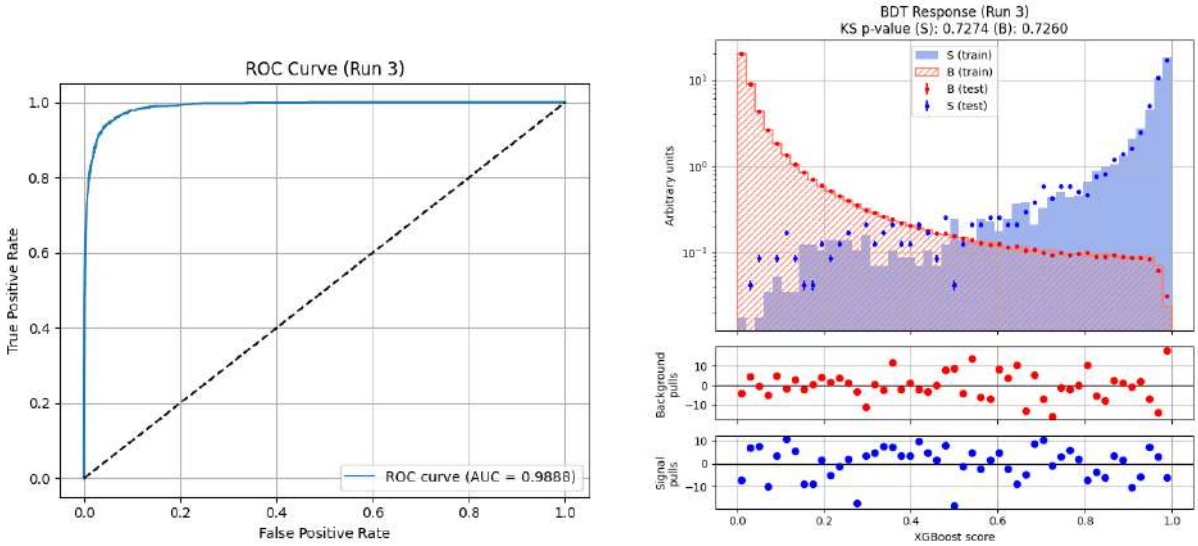
Performance is first assessed with receiver–operating characteristic (ROC) curves on independent test samples. For both runs the area under the curve (AUC) is close to unity demonstrating excellent separation between truth-matched signal and sideband background. To verify generalization, we compare train and test BDT-score distributions for both signal and background using Kolmogorov–Smirnov (KS) tests. This test measures the consistency between the two samples: high p-values indicate good agreement (no overtraining), while low p-values would suggest potential mismodeling or overfitting. The p-values remain comfortably high, with no significant mismatches between train and test sample.



(a) ROC curve of run 2

(b) BDT response of run 2

Figure 33: ROC curve and BDT response of run 2 dataset training



(a) ROC curve of run 3

(b) BDT response of run 3

Figure 34: ROC curve and BDT response of run 3 dataset training

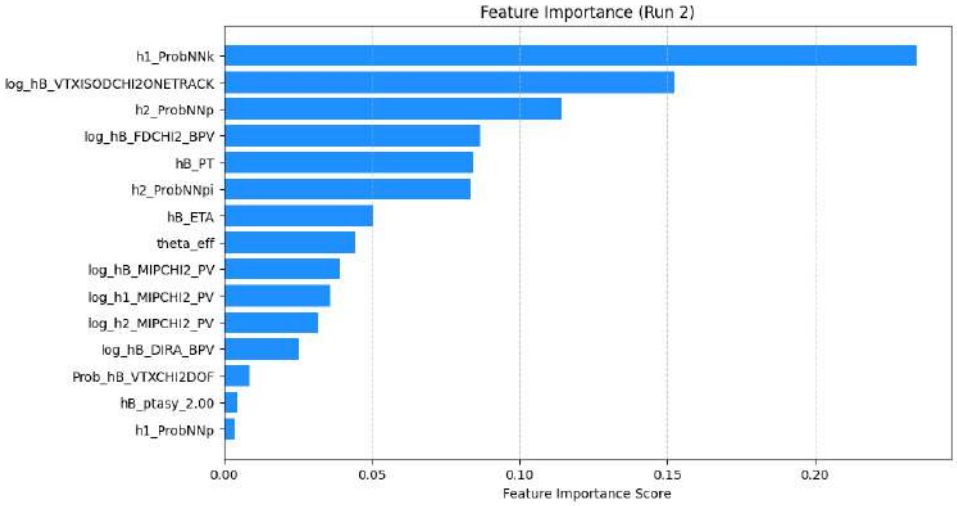


Figure 35: Feature importance of input variable in run2

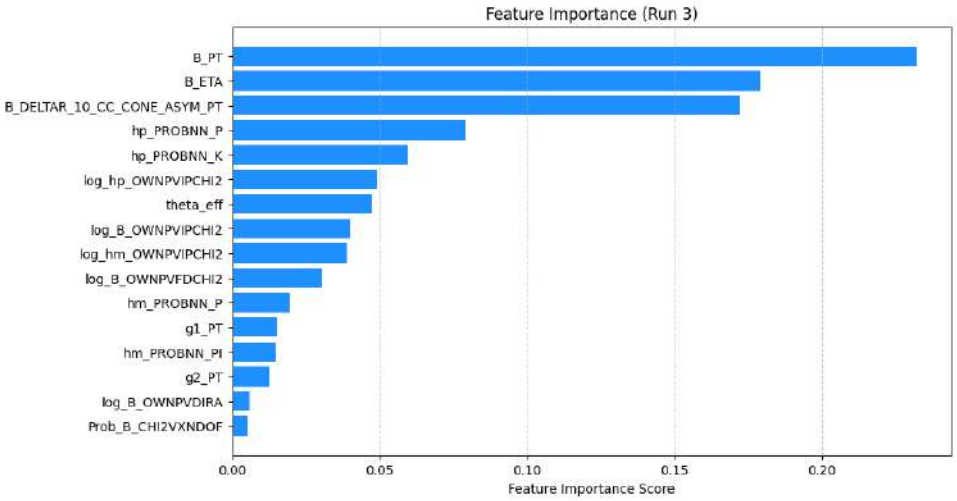


Figure 36: Feature importance of input variable in run3

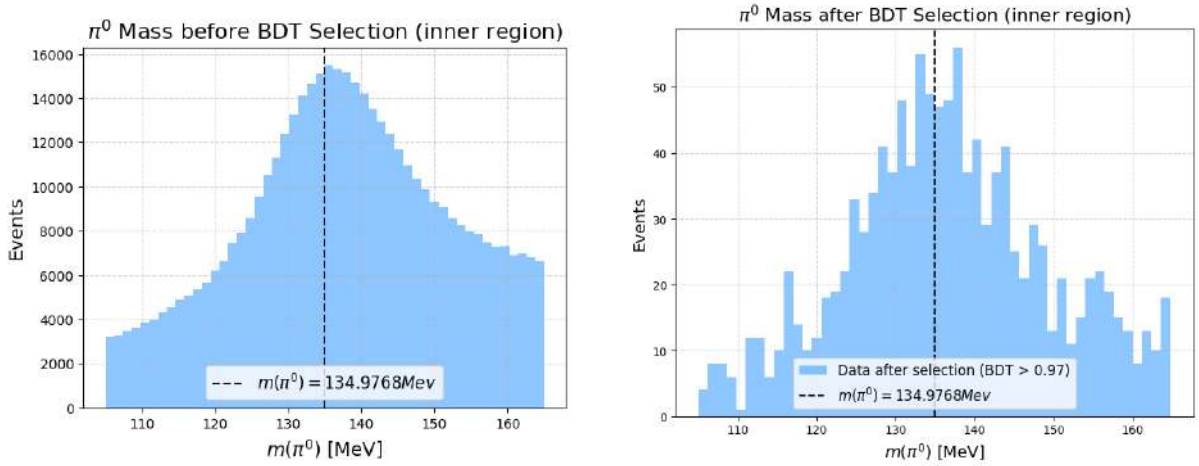
The boosted decision tree (BDT) classifier performance was assessed using ROC curves and the classifier output distributions. For Run 2, the ROC curve gives an AUC of 0.9919, demonstrating excellent separation between signal and background. The BDT response shows well-separated score distributions with good consistency between training and test samples. There is no evidence of overtraining, as demonstrated by the KS test values of 0.6937 for the signal and 0.8940 for the background, which are both safely above standard thresholds.

For Run 3, the classifier retains similarly strong performance, with an AUC of 0.9888. The separation between signal and background is still clear, although a slightly larger overlap can be seen compared to Run 2, which is consistent with the more challenging reconstruction conditions in Run 3. At low BDT scores, the behavior of the signal is affected by the limited size of the simulation sample, rather than a genuine modeling issue. This effect was explicitly checked and shown not to bias the training, as confirmed by the KS tests. The KS test values of 0.7224 for signal and 0.7260 for background again demonstrate good consistency between training and test samples. Overall, the results validate that the selected input variables and BDT configuration achieve effective suppression of background while maintaining stable performance and discrimination in both Run 2 and Run 3.

5 Reconstruction studies and Discussion

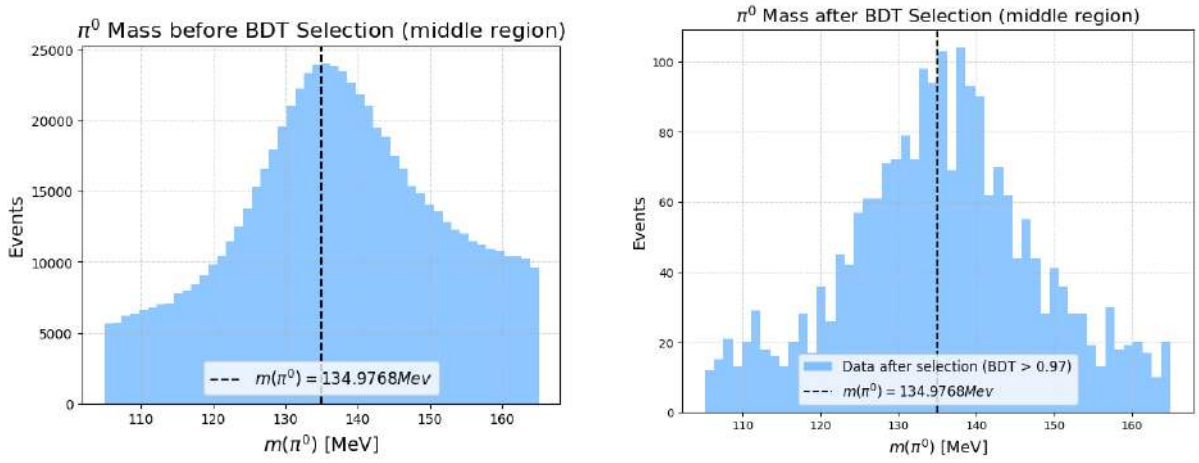
5.1 π^0 Reconstruction

Mass distributions before/after BDT and ECAL region dependence
Run 2:



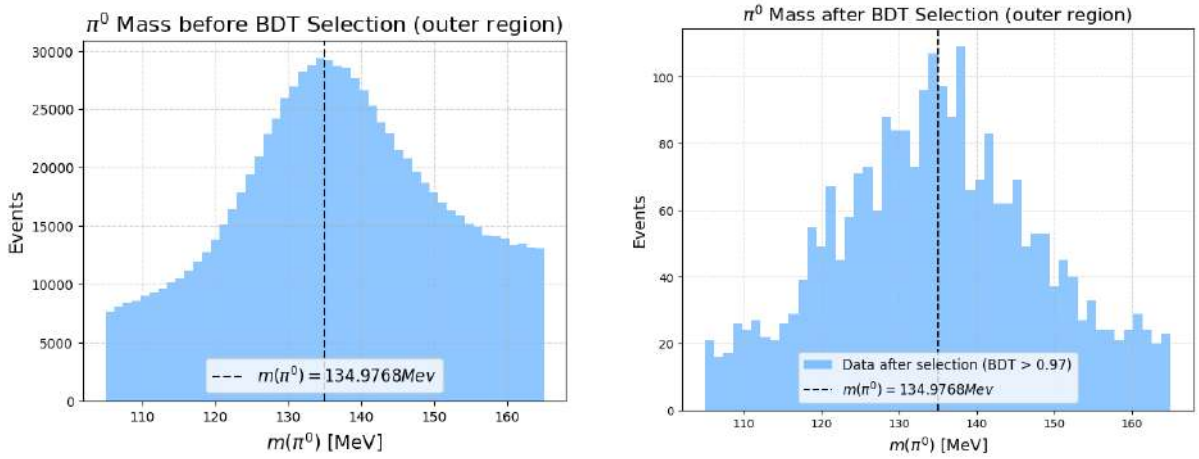
(a) (b)

Figure 37: Invariant mass distribution of neutral pions (π^0) in the ECAL inner region from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.97.



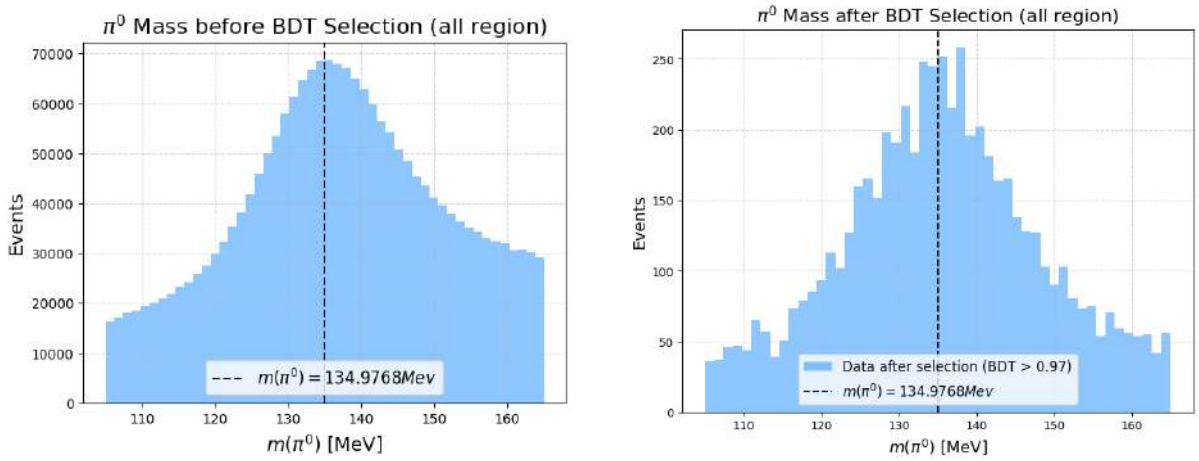
(a) (b)

Figure 38: Invariant mass distribution of neutral pions (π^0) in the ECAL middle region from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.97.



(a) (b)

Figure 39: Invariant mass distribution of neutral pions (π^0) in the ECAL outer region from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.97.



(a) ROC curve of run 2 (b) BDT response of run 2

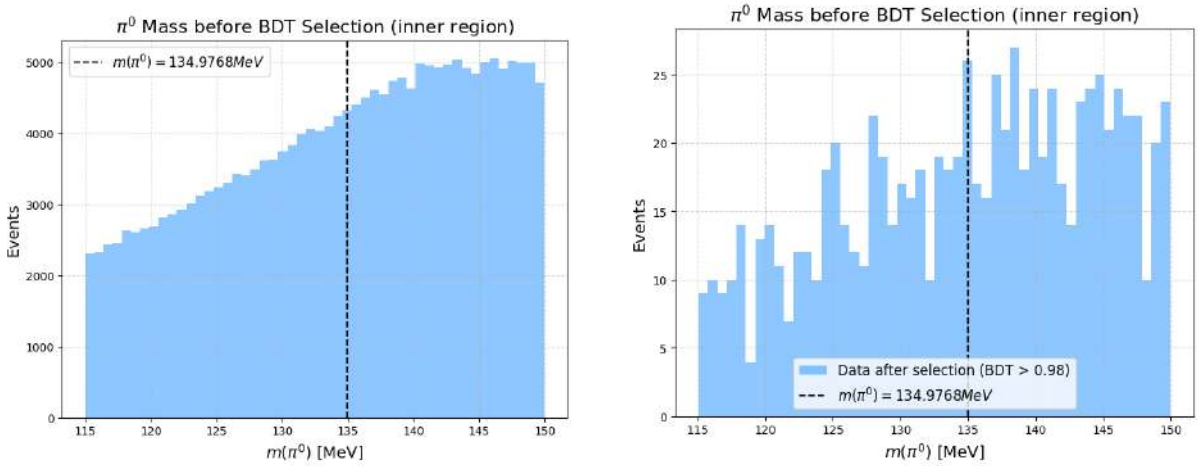
Figure 40: Invariant mass distribution of neutral pions (π^0) in the ECAL from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.97.

The invariant mass distributions of π^0 candidates in Run 2 show a well-defined peak around the PDG value of $m(\pi^0) \approx 134.98$ MeV/c^2, confirming the correct reconstruction of neutral pions across the detector regions. Before the application of the BDT selection, the spectra are dominated by combinatorial background from random photon pairs, producing broad tails in all regions. Nevertheless, the π^0 signal peak is clearly visible in the inner, middle, and outer calorimeter regions, and by eye the mean appears compatible with the expected π^0 mass across all three regions.

After applying the BDT selection, the background is significantly reduced, leading to a much cleaner signal peak in all regions. The inner region displays the sharpest mass distribution, consistent with its finer granularity and improved ability to resolve close photon pairs. The outer region, while still showing a visible peak, retains somewhat broader mass distributions due to coarser segmentation. The whole region of ECAL distribution illustrates the combined

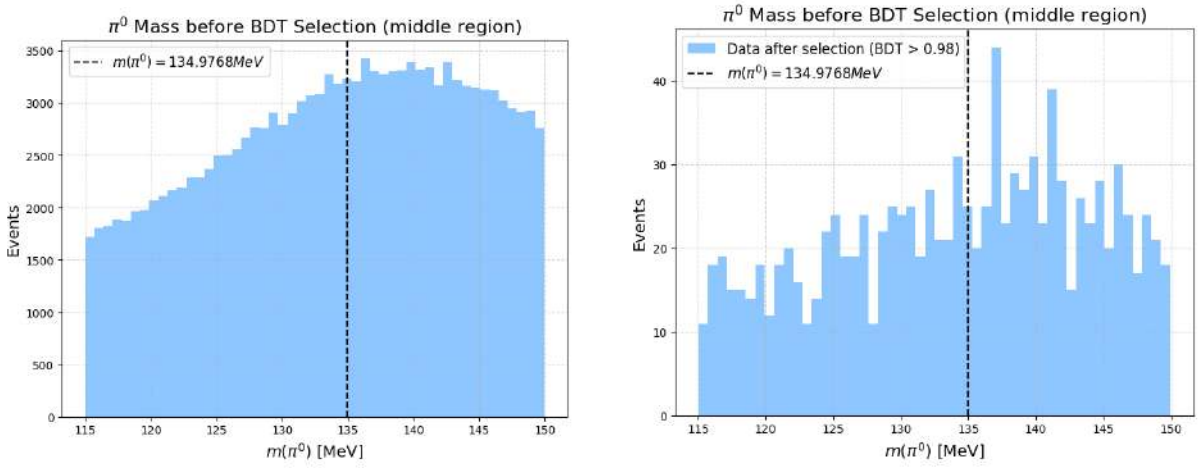
effect, maintaining consistency with the expected π^0 mass while demonstrating the effectiveness of the BDT in improving the signal-to-background ratio. Overall, these results confirm both the stability of the ECAL calibration in Run 2 and the crucial role of multivariate selection in enhancing neutral pion reconstruction.

Run 3:



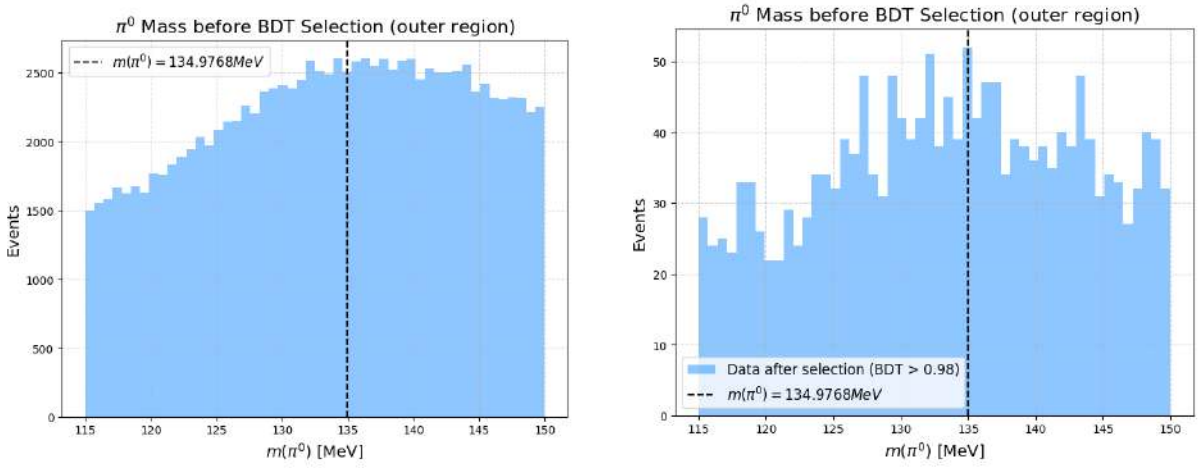
(a) (b)

Figure 41: Invariant mass distribution of neutral pions (π^0) in the ECAL inner region from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.



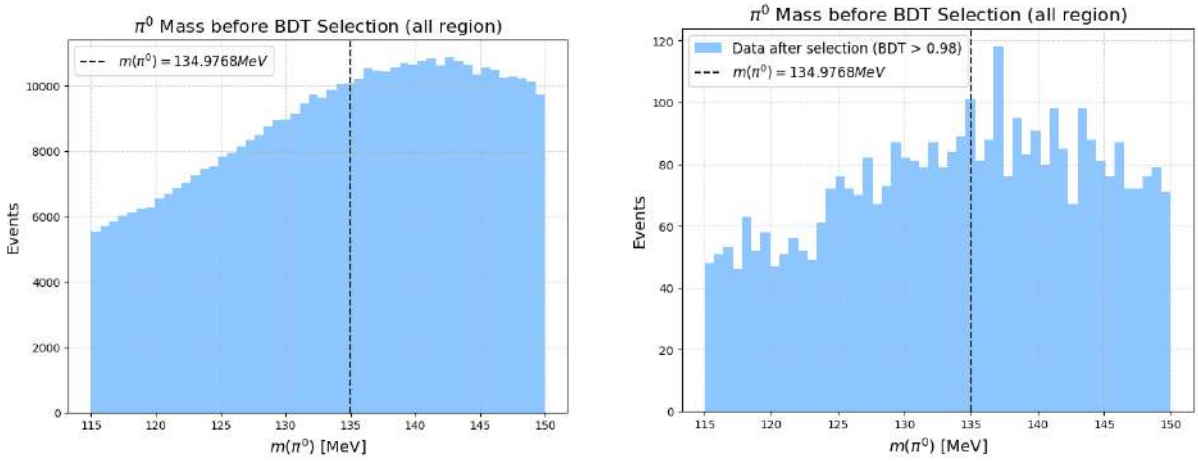
(a) (b)

Figure 42: Invariant mass distribution of neutral pions (π^0) in the ECAL middle region from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.



(a) (b)

Figure 43: Invariant mass distribution of neutral pions (π^0) in the ECAL outer region from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.



(a) (b)

Figure 44: Invariant mass distribution of neutral pions (π^0) in the ECAL from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.

The invariant mass distributions of π^0 candidates in Run 3 exhibit a less pronounced peak compared to Run 2. Before the BDT selection, the spectra are largely dominated by combinatorial background, and no clear signal peak is observed in any ECAL region. This behavior is consistent across the inner, middle, and outer regions. In addition to possible calibration shifts, higher detector occupancies, or modifications in the reconstruction algorithms, the reduced visibility of the π^0 signal in Run 3 is also influenced by the tighter π^0 mass window, which makes identification of the peak more difficult.

After applying the BDT selection, the background level is significantly suppressed. Nevertheless, the signal peak appears less well defined than in Run 2, and the peak tends to shift slightly upward, particularly in the inner region and in the inclusive all-region sample. These observations point to possible limitations in the current π^0 calibration of the ECAL and may also reflect mismodeling between data and simulation in Run 3. Taken together, the results highlight

the need for improved photon and π^0 calibrations, alongside region-dependent performance studies, to ensure accurate and reliable neutral pion reconstruction in the upgraded detector conditions.

5.2 π^0 Mass Fits Across Detector Regions under BDT Selections

Run 2:

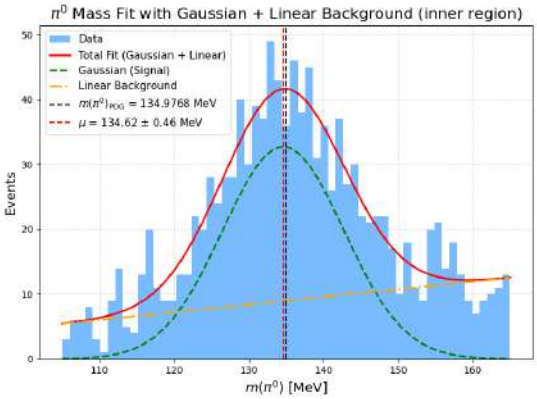


Figure 45: π^0 mass distribution in the ECAL inner region with Gaussian + linear fit.

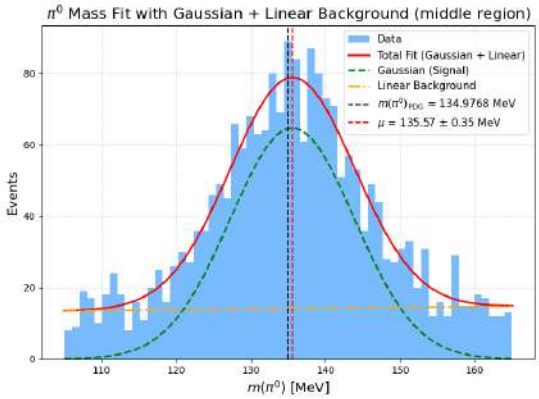


Figure 46: π^0 mass distribution in the ECAL middle region with Gaussian + linear fit.

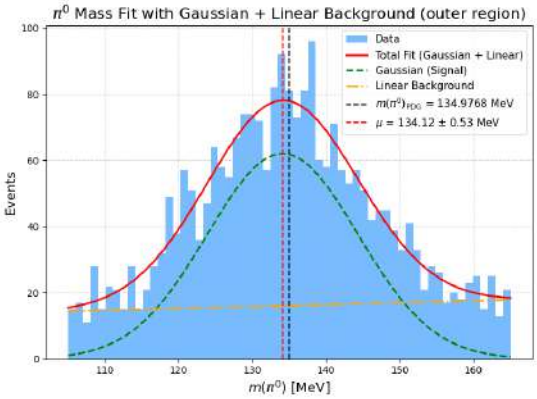


Figure 47: π^0 mass distribution in the ECAL outer region with Gaussian + linear fit.

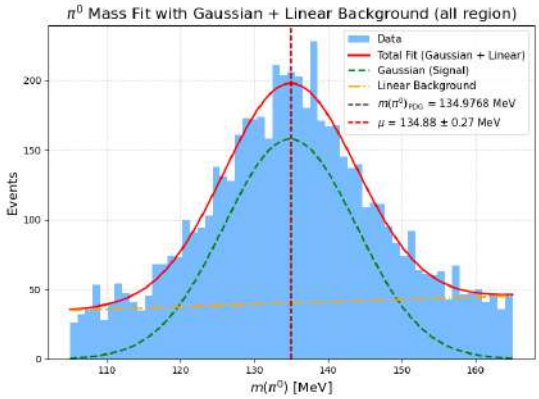


Figure 48: π^0 mass distribution in the ECAL with Gaussian + linear fit.

	Mean [MeV]	**Gaussian width [MeV]**
Inner region	134.618 ± 0.462	8.122 ± 0.556
middle region	135.568 ± 0.349	8.322 ± 0.424
Outer region	134.120 ± 0.534	10.220 ± 0.718
All region	134.880 ± 0.273	8.973 ± 0.342

Table 5: The table shows the mean (average π^0 mass) and the width of the fits in run2

Run 3:

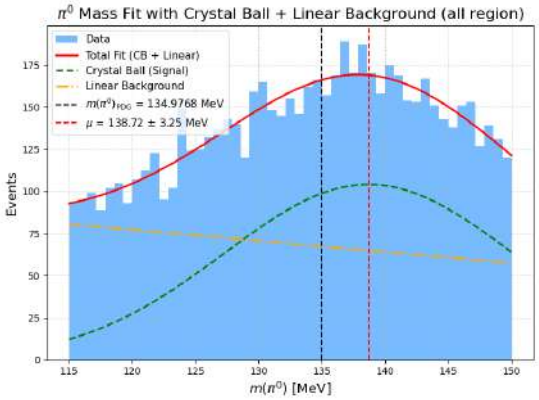


Figure 49: π^0 mass distribution in the ECAL with Crystall ball + linear fit.

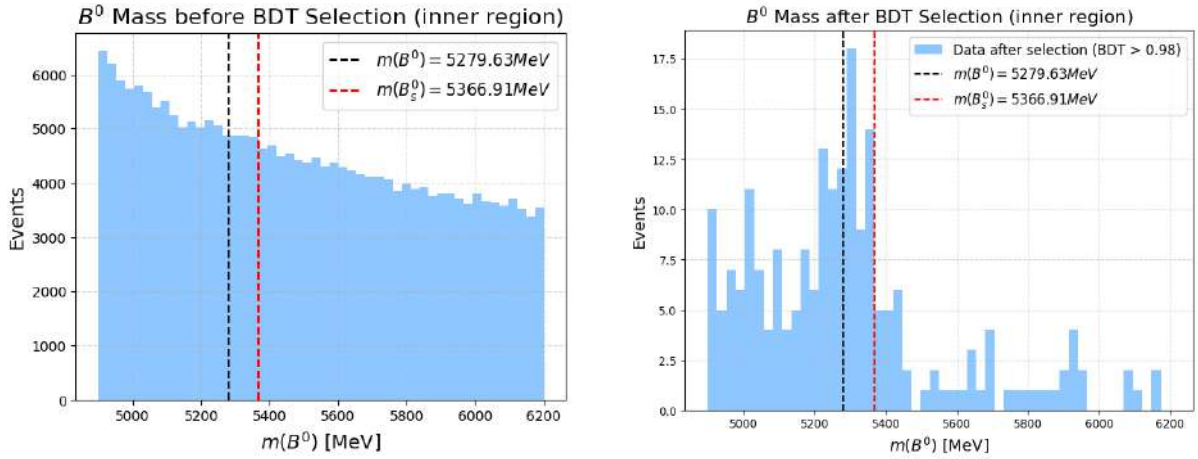
The fits to the π^0 mass distributions highlight a clear contrast between Run 2 and Run 3. In Run 2, the Gaussian plus linear background model provides stable fits across regions, with mean values close to the PDG mass of $m(\pi^0) = 134.98$ MeV/c^2. The fitted peaks are narrow and well defined, reflecting both good calorimeter calibration and reliable photon reconstruction.

By contrast, in Run3 the π^0 peak appears shifted upward ($m(\pi^0) \approx 138.72$ MeV/c^2) and significantly broader, as captured by the Crystal Ball plus linear background fit. The extracted mean deviates from the PDG value, and the uncertainty is larger compared to Run 2, suggesting reduced resolution and possible calibration mismodeling. In addition, the inconsistency observed in Run 3 may be partly influenced by the narrower mass window used in the fit compared to Run 2, which can distort the fitted parameters. However, the fit provides only an approximate estimate of the mean due to the limited π^0 mass window, therefore, the result should not be interpreted as a precise determination. In summary, these differences underline the need for improved photon response studies and dedicated π^0 calibration to ensure reliable performance in the upgraded Run 3 environment.

6 B Mass Reconstruction

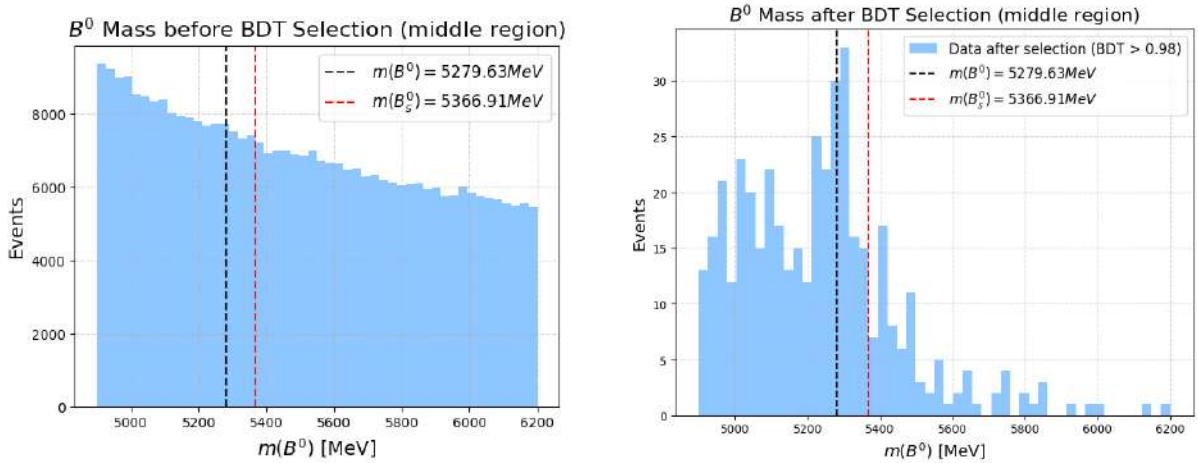
6.1 Mass distributions before/after BDT and ECAL region dependence

Run 2:



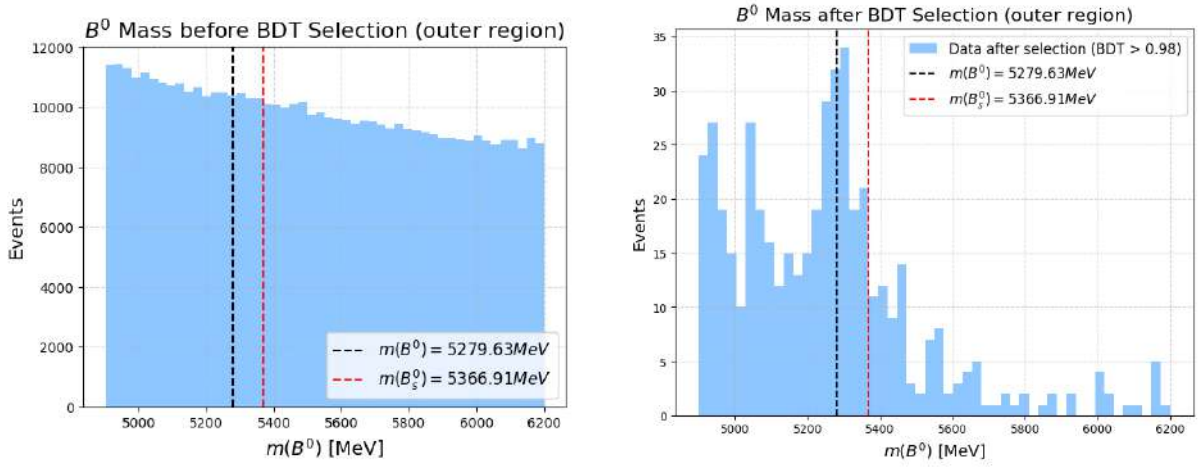
(a) (b)

Figure 50: Invariant mass distribution of B^0 meson in the ECAL inner region from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.



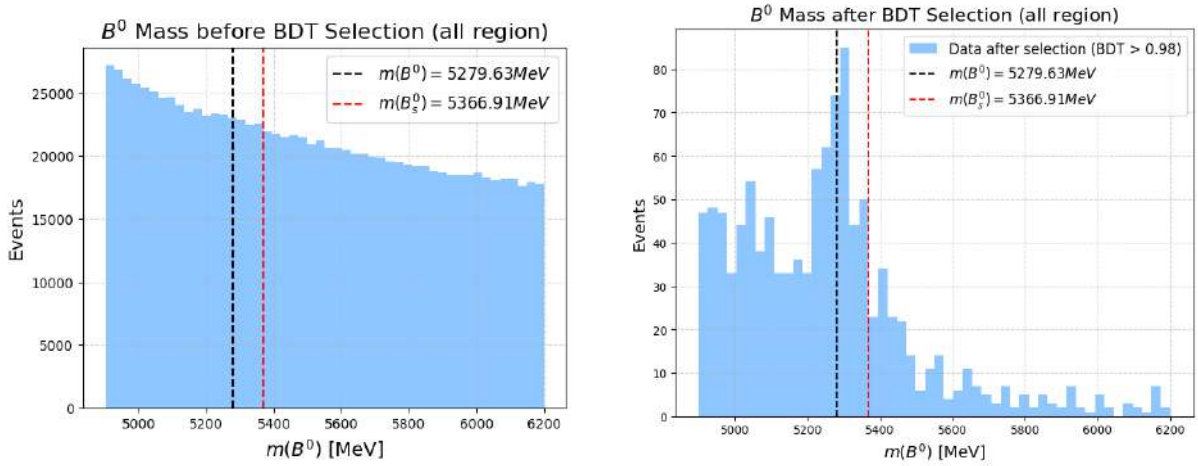
(a) (b)

Figure 51: Invariant mass distribution of B^0 meson in the ECAL middle region from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.



(a) (b)

Figure 52: Invariant mass distribution of B^0 meson in the ECAL outer region from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.



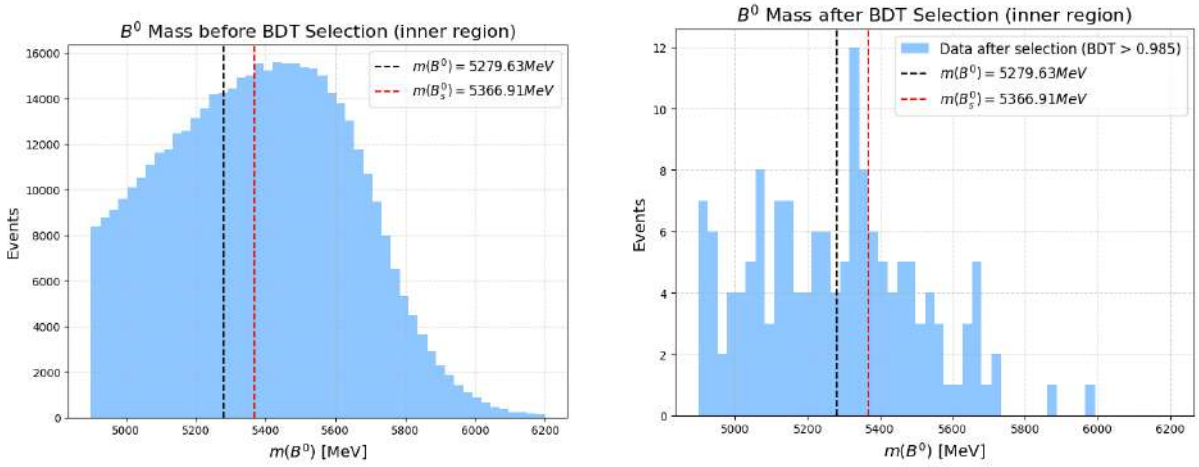
(a) (b)

Figure 53: Invariant mass distribution of B^0 meson in the ECAL from a Run 2 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.98.

The invariant mass distributions of B^0 candidates in Run 2 show a clear improvement after the application of the BDT selection. Before the selection, the spectra across all ECAL regions are dominated by combinatorial background. In the vicinity of the nominal B^0 mass of 5279.6 MeV/c^2, the distributions exhibit only a smeared excess of events rather than a distinct, narrow peak. This indicates that genuine B^0 decays are present but obscured by background contributions and limited mass resolution.

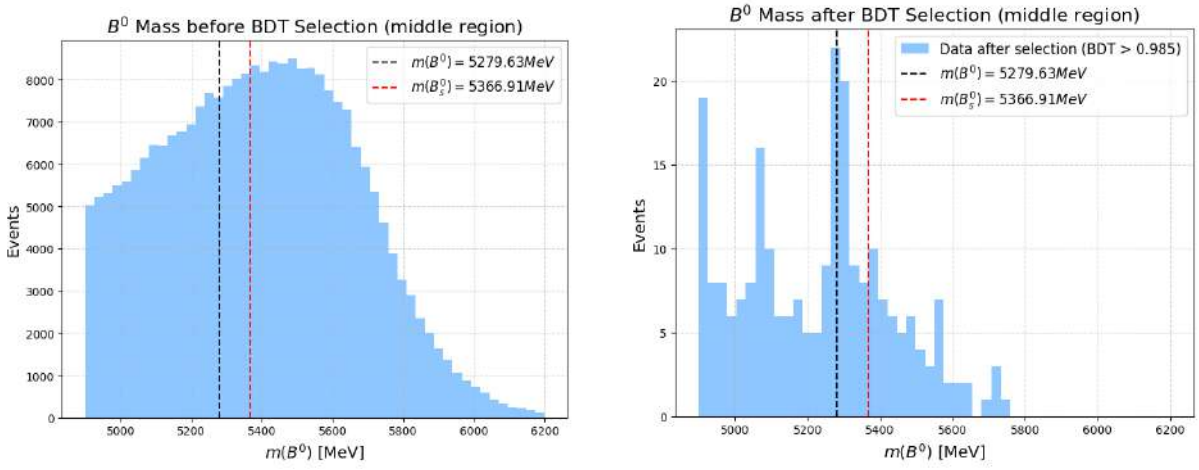
After applying the BDT cut, the background level is significantly reduced, and a visible signal emerges around the expected B^0 mass. This behavior is consistent across the inner, middle, outer, and inclusive ECAL regions, although the statistical precision remains limited due to the relatively small sample size after selection. The improvement demonstrates the discriminating power of the BDT, which efficiently suppresses random combinations while preserving genuine signal candidates. These results confirm the validity of the selection strategy and provide a basis for subsequent measurements of signal yields and branching fractions.

Run 3:



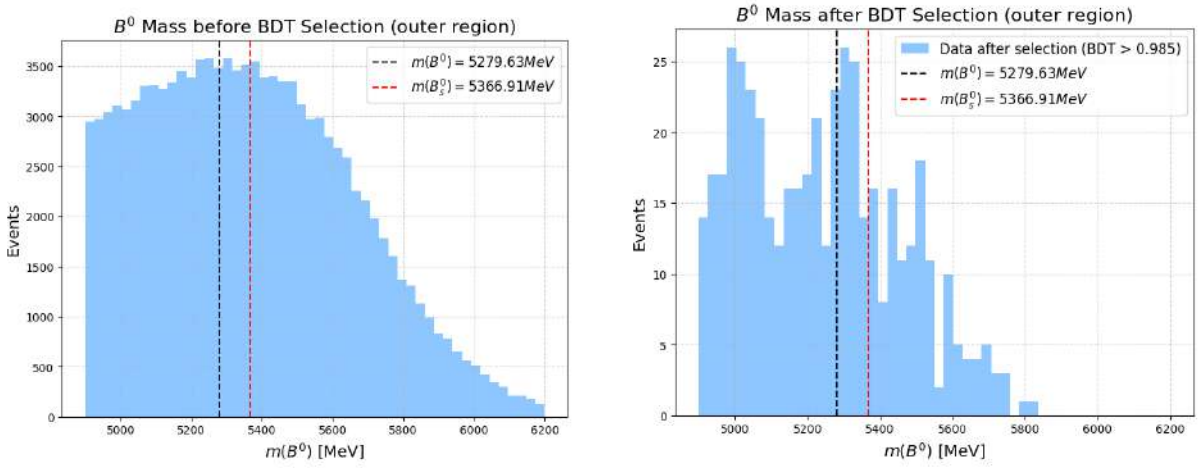
(a) (b)

Figure 54: Invariant mass distribution of B^0 meson in the ECAL inner region from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.985.



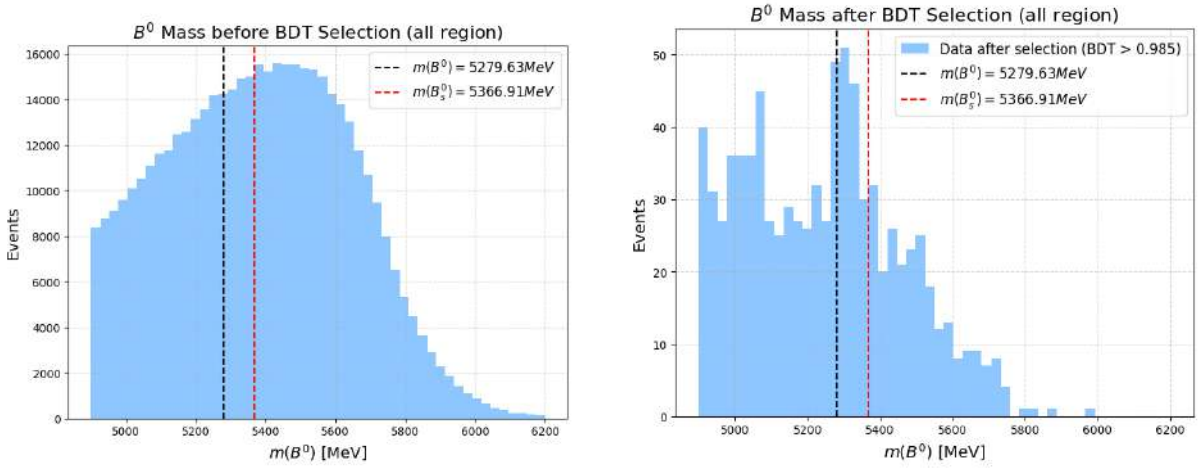
(a) (b)

Figure 55: Invariant mass distribution of B^0 meson in the ECAL middle region from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.985.



(a) (b)

Figure 56: Invariant mass distribution of B^0 meson in the ECAL outer region from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.985.



(a) (b)

Figure 57: Invariant mass distribution of B^0 meson in the ECAL from a Run 3 data. Plot (a) shows the distribution before BDT selection, and plot (b) shows the same distribution after a BDT selection with a cut of BDT > 0.985.

The invariant mass distributions of B^0 candidates in Run 3 exhibit a clear signal region, but the separation between signal and background is noticeably weaker compared to Run 2. Before the BDT selection, the spectra are dominated by combinatorial background, with only a modest excess around the expected B^0 mass of 5279.6 MeV/c^2. After applying the BDT cut,which was selected not by optimising a formal figure of merit but by comparing several threshold values and inspecting their effect on the reconstructed signal peak, the background is substantially reduced and a peak emerges close to the nominal mass in all ECAL regions. However, the signal remains broader and less well isolated than in Run 2.

This reduced separation can be attributed in part to the performance of π^0 reconstruction in Run 3. As shown in the π^0 studies, the reconstructed π^0 mass peak in Run 3 is shifted relative to Run 2. Since the $B^0 \to K^+\rho^-$ channel involves a neutral pion in the final state, the poorer π^0 resolution propagates directly to the reconstructed B^0 candidate mass. Consequently, the Run 3 results emphasize the critical dependence of B^0 meson mass resolution on neutral pion performance, underlining the need for improved calibration and reconstruction of photon final states in the upgraded detector.

7 Conclusion and Outlook

In this work, a study of neutral pion reconstruction in the decay channels relevant to B^0 meson analyses has been presented using LHCb Run 2 and Run 3 data. The results demonstrate that Run 2 provides a well-defined π^0 mass peak consistent with the PDG value, leading to stable reconstruction of both π^0 and B^0 candidates across ECAL regions. In contrast, Run 3 shows broader and shifted mass distributions, both for π^0 and B^0, reflecting challenges in neutral pion reconstruction under the upgraded detector conditions. These effects highlight the critical role of photon calibration and cluster reconstruction in achieving reliable performance in analyses involving neutral final states.

Looking forward, the immediate next step is to refine the calibration of π^0 reconstruction in Run 3, including region-dependent corrections and improved modeling of photon response. Furthermore, precise determination of the branching fraction for $B_s^0 \to K^-\pi^+\pi^0$ will require careful modeling of both B^0 and B_s^0 signal and background components, as well as efficiency corrections for the selection criteria, together with particle identification (PID) calibrations to align kaon/pion efficiencies and misidentification rates between data and simulation.

8 Appendices

Particle ID and BKGCAT interpretation

Particle ID [7]

Particle ID	Particle
321	K^+
211	π^+
111	π^0
22	γ
511	B^0

BKGCAT

- 0: Signal
- 50: Low-mass background

opening angle and theta_eff definition

θ_{eff} :

$$\theta_{eff} = \theta_{\gamma_1\gamma_2} \times \frac{r}{cellsize}$$

where
$\theta_{\gamma_1\gamma_2}$: opening angle, defines as

$$\theta_{\gamma_1\gamma_2} = arccos\frac{(g1\_PX * g2\_PX + g1\_PY * g2\_PY + g1\_PZ * g2\_PZ)}{(g1\_P * g2\_P)}$$

r defines as

$$\frac{12.51}{cos(arctan((\sqrt{((pi0\_PX * pi0\_PX) + (pi0\_PY * pi0\_PY))}/pi0\_PZ))}$$

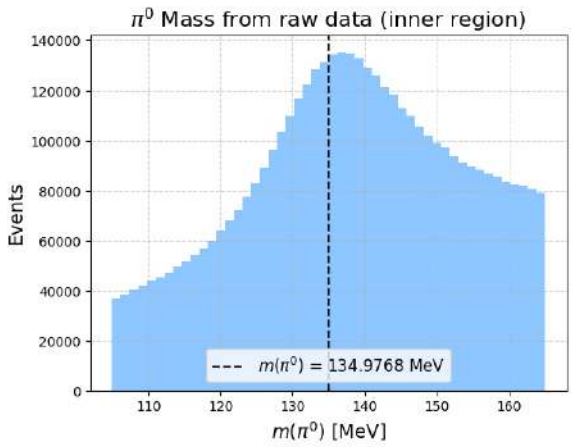
and definition of cell size per region is

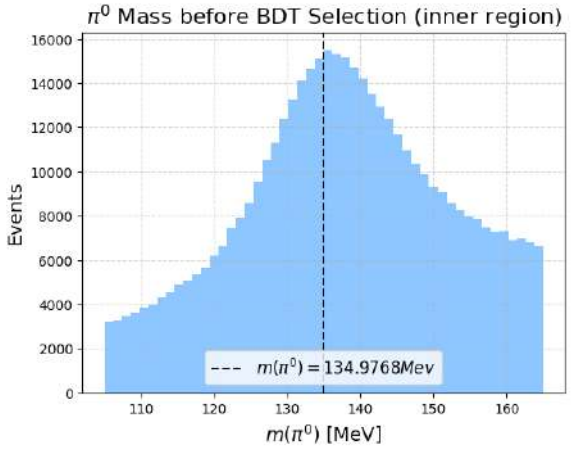
- cellsize_outer = 0.12 m
- cellsize_middle = 0.06 m
- cellsize_inner = 0.04 m

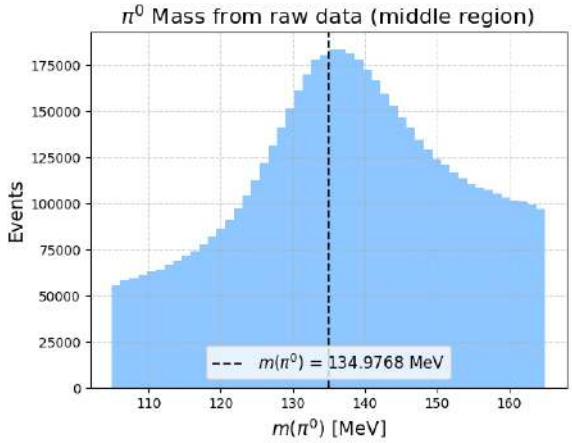
Mass distribution after event selection process

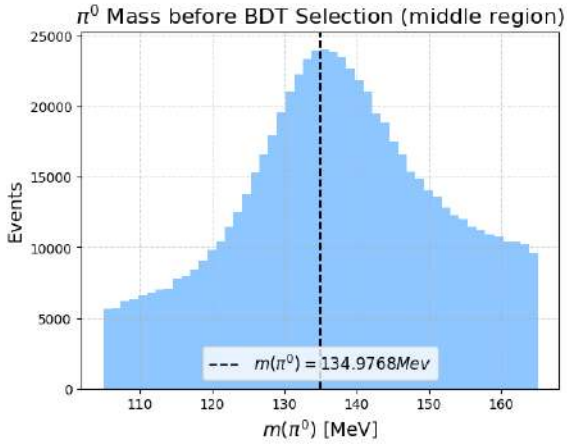
Run 2

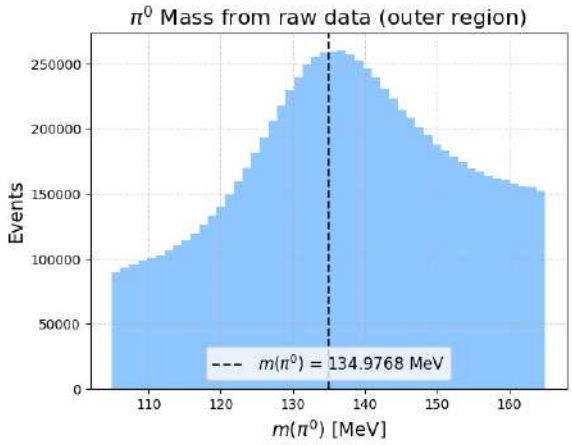
π^0 mass distribution : Comparison of the π^0 mass distribution between the raw data (left) and the data after applying event selection (right).

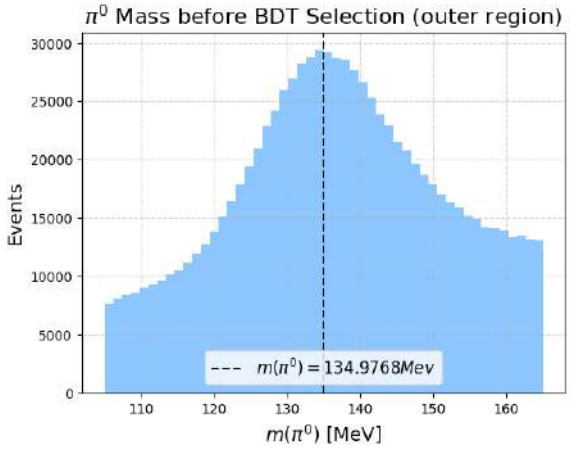


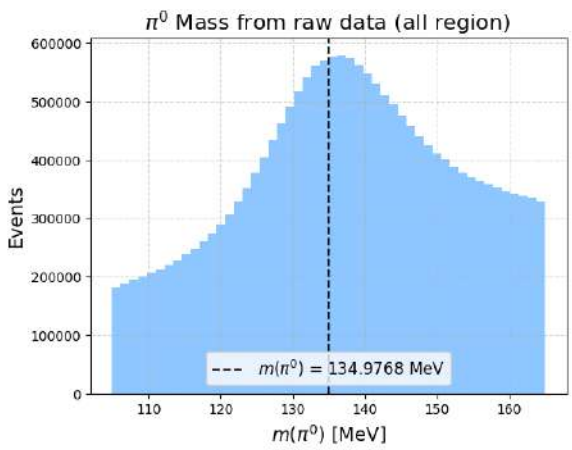


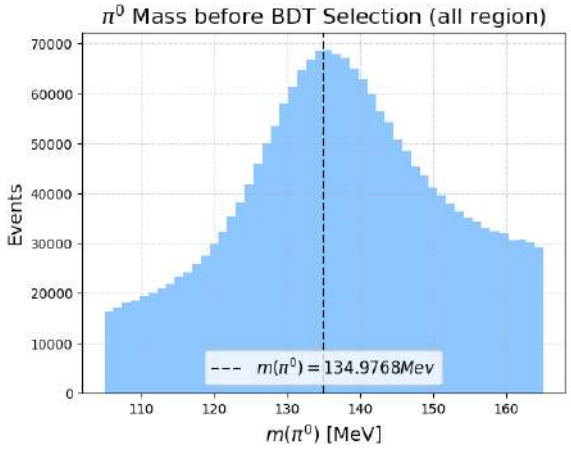




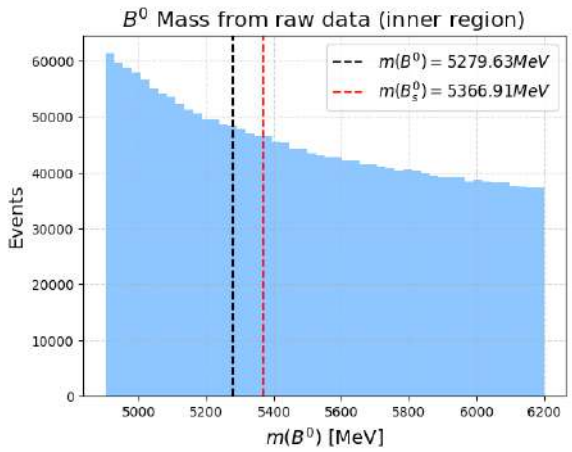


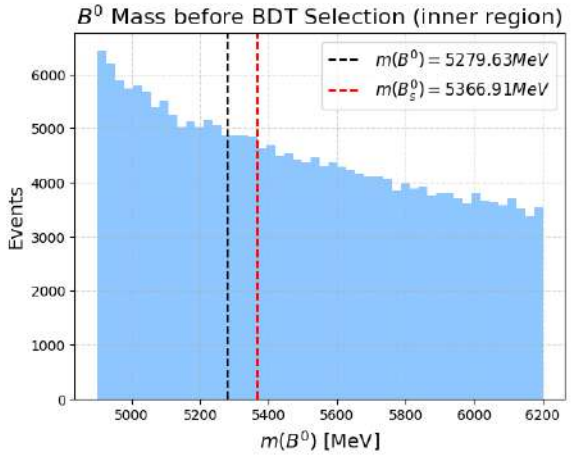


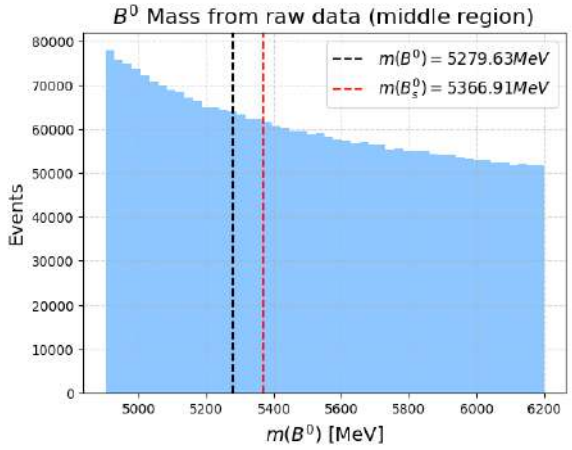


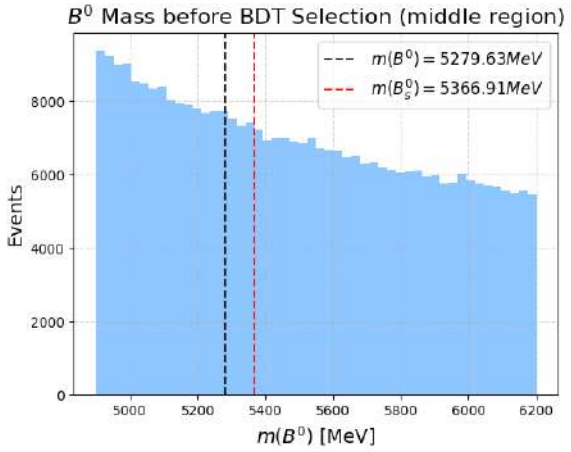


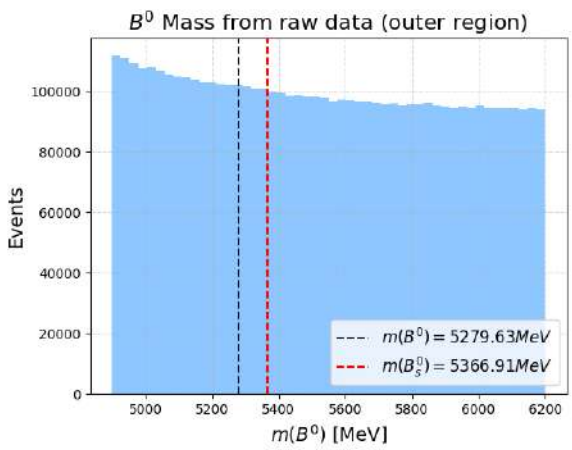
B mass distribution : Comparison of the B^0 mass distribution between the raw data (left) and the data after applying event selection (right).

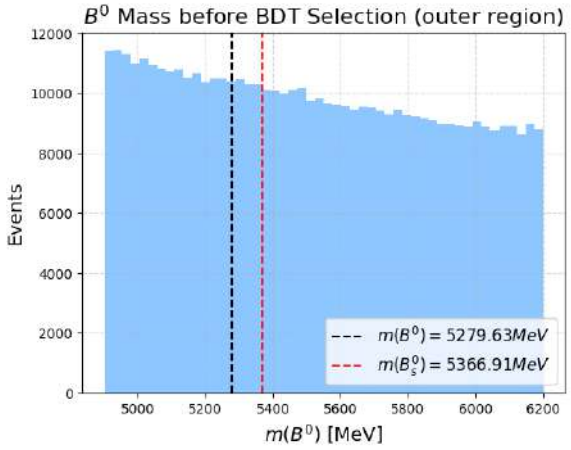


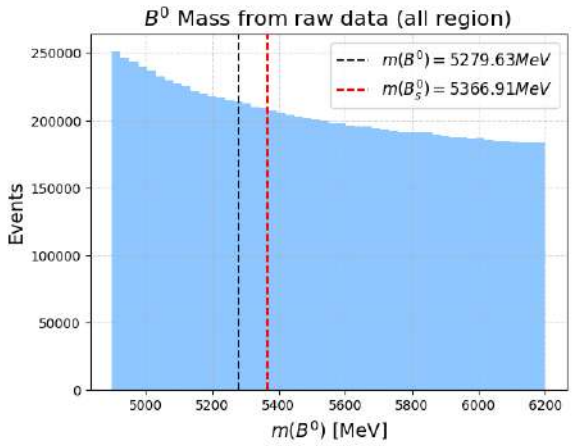


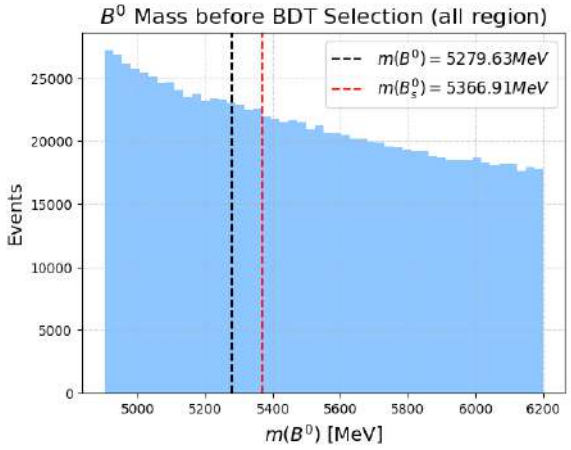






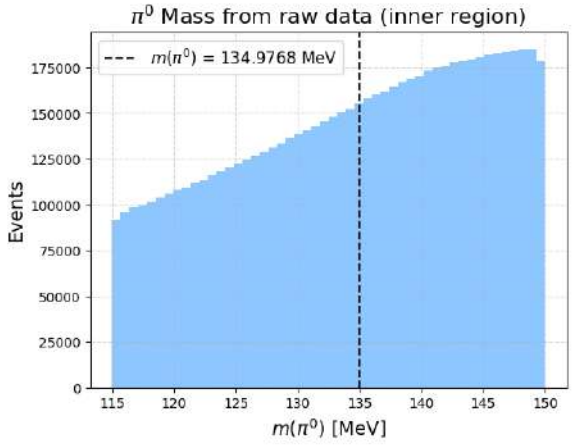


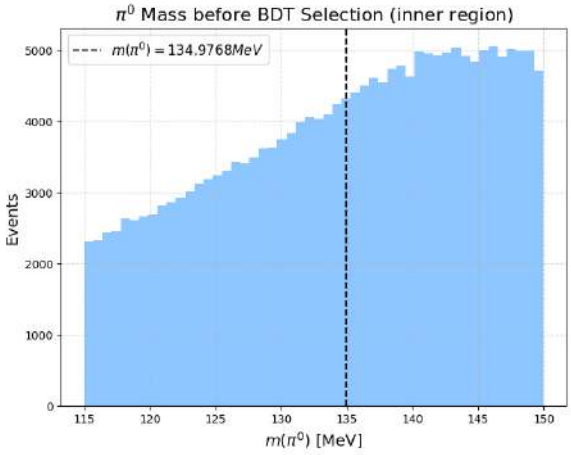


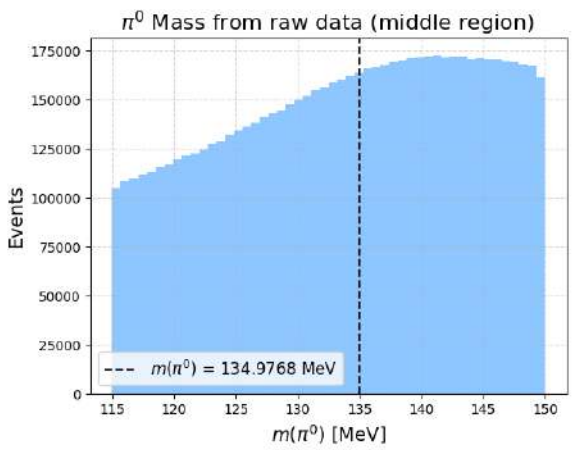


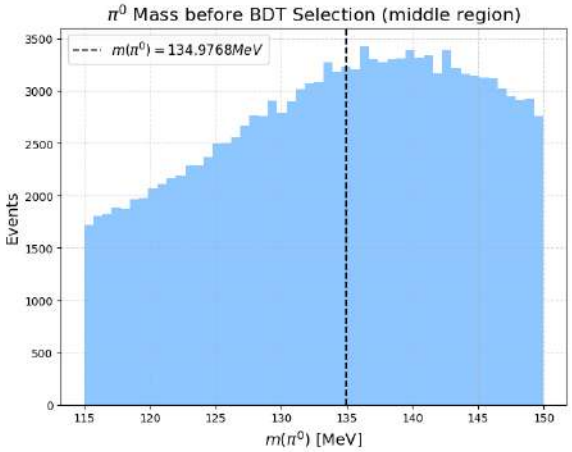
Run 3

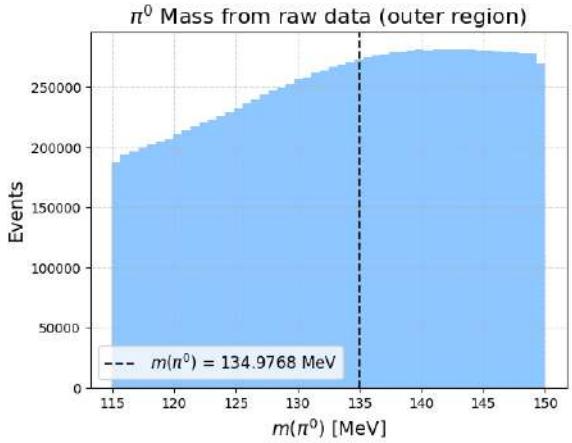
π^0 **mass distribution :** Comparison of the π^0 mass distribution between the raw data (left) and the data after applying event selection (right).

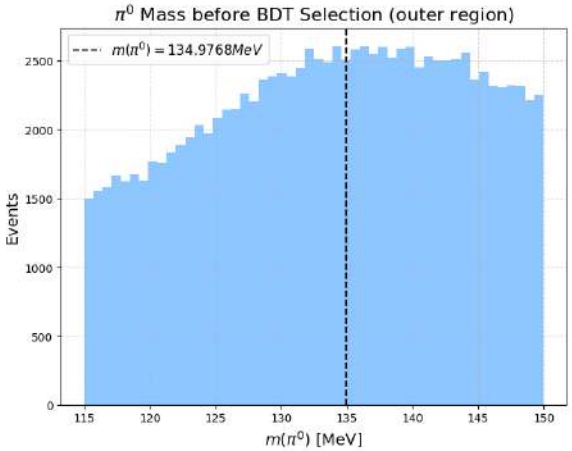


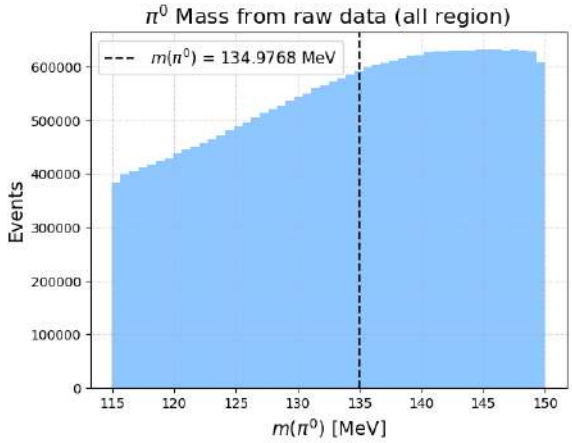


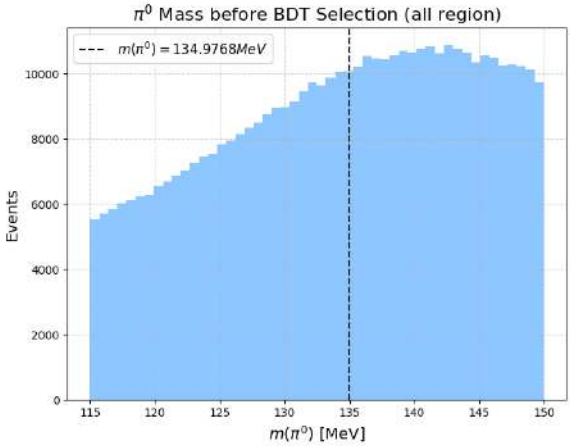

π^0 Mass from raw data (middle region)
Events
$m(\pi^0)$ [MeV]
$m(\pi^0)$ = 134.9768 MeV


π^0 Mass before BDT Selection (middle region)
Events
$m(\pi^0)$ [MeV]
$m(\pi^0) = 134.9768MeV$

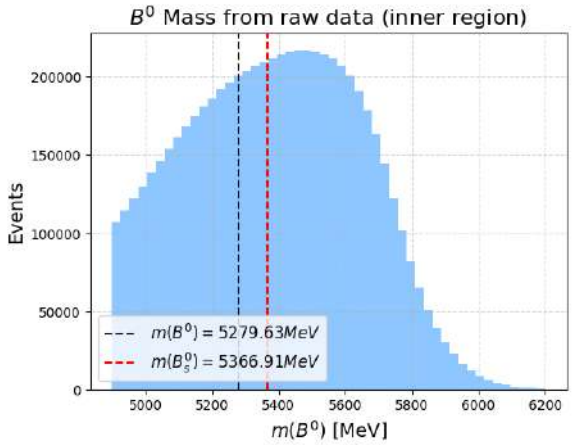

π^0 Mass from raw data (outer region)
Events
$m(\pi^0)$ [MeV]
$m(\pi^0)$ = 134.9768 MeV

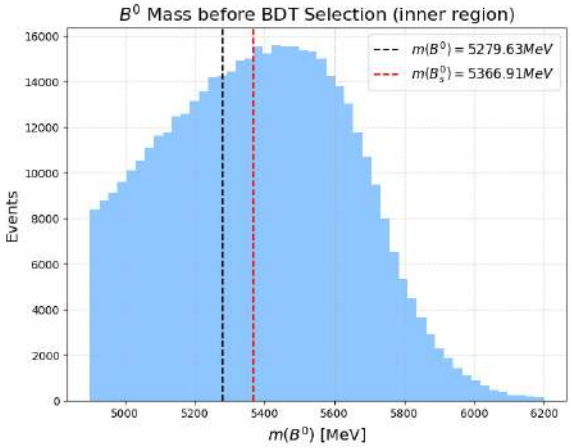

π^0 Mass before BDT Selection (outer region)
Events
$m(\pi^0)$ [MeV]
$m(\pi^0) = 134.9768MeV$

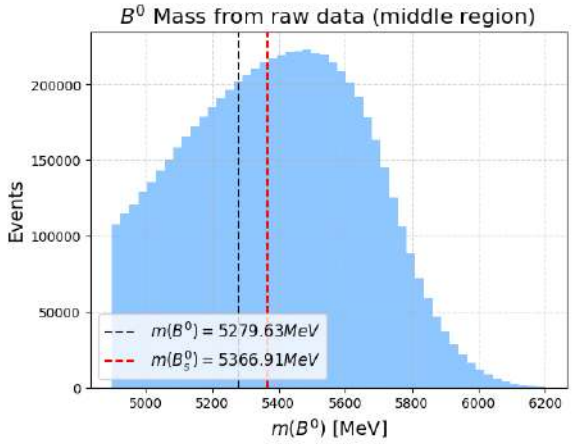

π^0 Mass from raw data (all region)
Events
$m(\pi^0)$ [MeV]
$m(\pi^0)$ = 134.9768 MeV

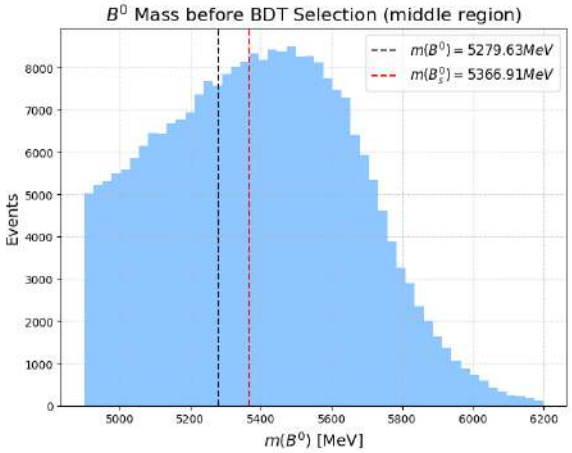

π^0 Mass before BDT Selection (all region)
Events
$m(\pi^0)$ [MeV]
$m(\pi^0) = 134.9768MeV$

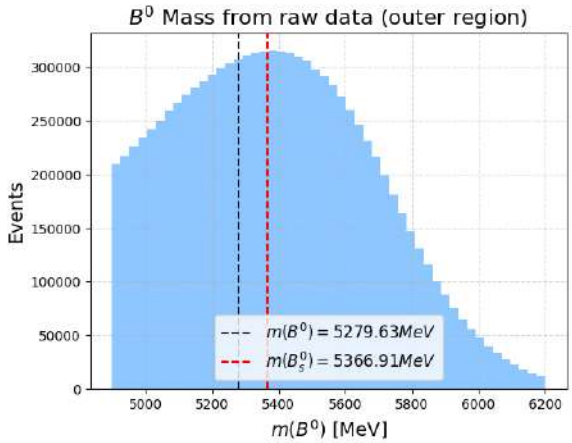
B mass distribution : Comparison of the B^0 mass distribution between the raw data (left) and the data after applying event selection (right).

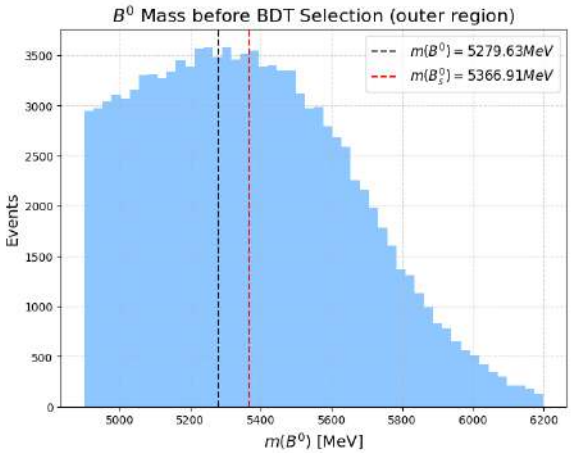


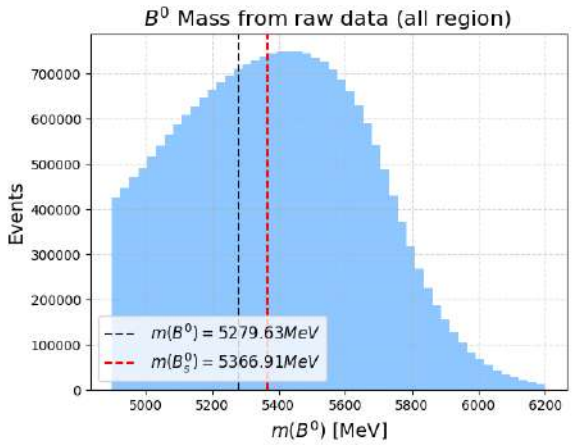


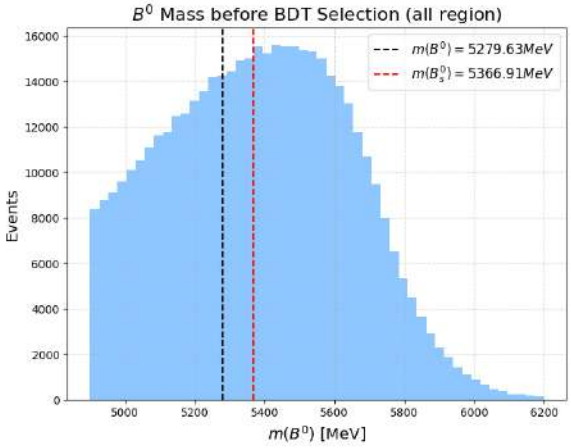












References

[1] J. J. Saborido Silva. *CP* violation in the *B* system at LHCb. In *49th Rencontres de Moriond on QCD and High Energy Interactions*, pages 85–90, 2014.

[2] O. Deschamps, F. P. Machefert, M. H. Schune, G. Pakhlova, and I. Belyaev. Photon and neutral pion reconstruction. 9 2003.

[3] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024.

[4] C. Abellán Beteta et al. Calibration and performance of the lhcb calorimeters in run 1 and 2 at the lhc, 2020.

[5] Wouter D. Hulsbergen. Decay chain fitting with a kalman filter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 552(3):566–575, November 2005.

[6] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machinery.

[7] L. Garren, C.-J. Lin, Sergio Navas, P. Richardson, T. Sjóstrand, et al. Monte carlo particle numbering scheme.

Diary

TUE 10th June

This morning, I attended the welcome session for the Summer Student Program 2025. The session covered essential information and guidelines, including important rules, proper conduct, as well as details on medical services and insurance. After the session, I went to Building 2 to meet my supervisor, Jordy Butter. He gave me a brief overview of my project and we had lunch together. In the afternoon, I began trying to connect to the CERN server and was assigned to start learning about Bash, which is used for server connection and basic command-line operations.

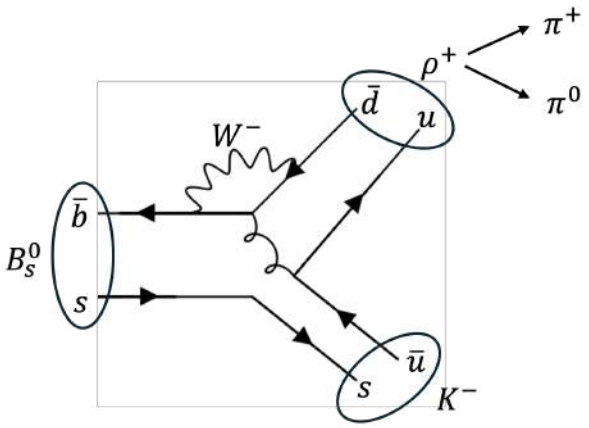


Figure 74: Feynman diagram of B_s^0 decay

WED 11st June

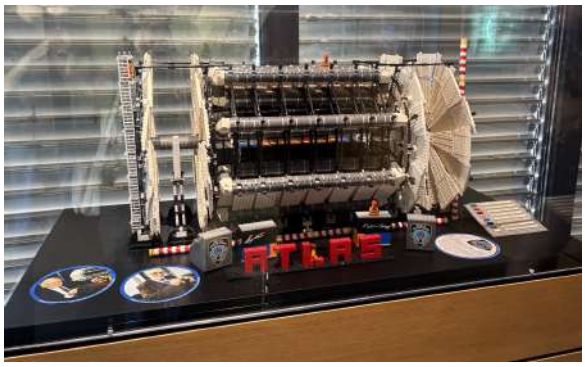
Today, I was still busy with accessing LXPLUS, continuing to familiarize myself with the environment and troubleshooting some issues that came up. I also try to use CodIMD to write my progress report. It was a good opportunity to get used to the system and improve my workflow. In the afternoon, I decided to take a break and experience something new: I took the shuttle bus from the CERN dormitory to the Carrefour for the first time. It was a convenient ride, and I was able to pick up a few items I needed. The experience felt like a small adventure, navigating the CERN campus and exploring a bit more of the facilities. Overall, it was a productive day with a nice mix of work and a bit of exploration.

THU 12nd June
Today, I went to building 73 to get my office key. I got lost while I was walking to such a building. I spent around an hour instead of just 15 mins as usual, but at least I can explore new ways myself. After that, I went back to the office, continuing my previous work and reading some relevant research papers to understand more about this work. In the evening, I decided to try cooking for myself with some ingredients I bought from the carrefour. It was a nice break from the routine, and I enjoyed the process of preparing a meal. The experience of cooking in a new place made the day even more fulfilling.

Figure 75: The corridor in front of my office.

FRI 13rd June
Today, I didn't do anything particularly new. I continued to review and study how to use zfit. I spent the day familiarizing myself further with its functionalities and exploring its features. While there weren't any major breakthroughs today, I made steady progress in understanding how to apply it for fitting data, which is crucial for the ongoing analysis. Overall, it was a day of reinforcing what I've learned so far with zfit.

MON 16th June
In the morning, I presented my progress report from last week to my supervisor here. After that, I worked on constructing the Dalitz plot as part of my ongoing project. In the afternoon, I joined a site visit to both the ATLAS experiment and the synchrocyclotron, CERN's first accelerator. It was a very interesting visit that provided valuable and detailed insights into the ATLAS detector and the synchrotron system. I found it to be an informative and engaging experience.



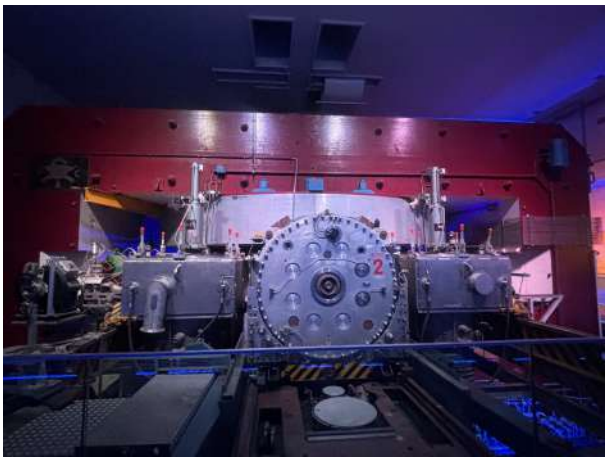


Figure 77: ATLAS experiment and ATLAS detector (above) and the synchrocyclotron (below)

TUE 17th June
Today, I explored data from real collisions and compared it with MC simulation data to analyze the differences between the two. I observed that the real data contained a significant amount of background, which made it more challenging to distinguish the signal from noise. This comparison provided valuable insights into the nature of real experimental data and highlighted the importance of improving methods for background suppression in the analysis. It was an interesting exercise that helped deepen my understanding of the complexities involved in real-world data processing.

WED 18th June
Today, I visited the antimatter lab, which was very fascinating. During the visit, there was a presentation about various detectors, and we had a discussion exchanging views on the topics of dark matter and antimatter, particularly those aspects that the Standard Model cannot explain.

Figure 78: Building of antimatter experiment

THU 19th June
Additionally, I studied the Dalitz plot, which is a graphical representation used to analyze the kinematics of three-particle decay processes. The Dalitz plot helps visualize the energy distribution between the particles and is commonly used to identify resonances or patterns in particle decays. I plan to use this plot to further analyze decay channels and potentially extract meaningful insights about particle interactions in the simulation data.

FRI 20th June
Today, my supervisor wasn't around, so I had to work alone. There weren't many tasks to tackle, so I spent the day making some adjustments to the data analysis, addressing a few issues I had encountered earlier in the week. After completing the tasks, I decided to reward myself for the hard work throughout the week. I took a walk at the botanical garden, which was a relaxing break. Later, I treated myself to a delicious meal of green curry and mango sticky rice at a Thai restaurant. It was a nice way to unwind and reflect on the week's progress.

Figure 79: Mango sticky rice at a Thai restaurant

MON 23rd June
Today, my supervisor is still on vacation, so I spent the day reading more papers related to my work to deepen my understanding. I also worked on updating my progress report slides based on the previous meeting, adding more details about my current work. Additionally, I recreated the Dalitz plot to refine my analysis and better prepare for tomorrow's presentation. I focused on ensuring that I could clearly explain my progress and the insights I've gained so far.

TUE 24th June
Today, I reported my progress to my supervisor and the co-advisor, sharing the updates on the simulation analysis for Run 2. They provided feedback on areas that need improvement in the analysis. Based on their suggestions, I was assigned the task of applying the same analysis approach to both the merged and resolved files from the experimental data. This will allow me to compare and refine the methods, ensuring consistency in analyzing the real experimental data alongside the simulation.

WED 25th June
Today, I focused on refining the feature selection process for the machine learning training. After reviewing the initial results, I identified some features that may not contribute effectively to the classification and could even introduce noise into the model. I spent the day experimenting with different combinations of features to see how they affected the performance. It was a challenging but insightful task, laying the groundwork for more robust training in the coming days.

THU 26th June
Today, I joined a site visit to the LHCb Experiment as part of the summer student program. The visit took place on the French side of CERN, where I had the opportunity to learn about the work of researchers involved in the LHCb experiment. We were introduced to the functions of the detector systems and how various quantities from the sub-detectors are recorded into ROOT files, which are later used for data analysis. One of the highlights of the visit was going underground to see the LHCb detector in person—it was an awe-inspiring experience to stand so close to such a sophisticated piece of scientific equipment. On my way back to the dormitory, I walked along a beautiful, nature-filled path. It was a peaceful and refreshing walk, and to my delight, I saw Swiss cows up close for the first time.

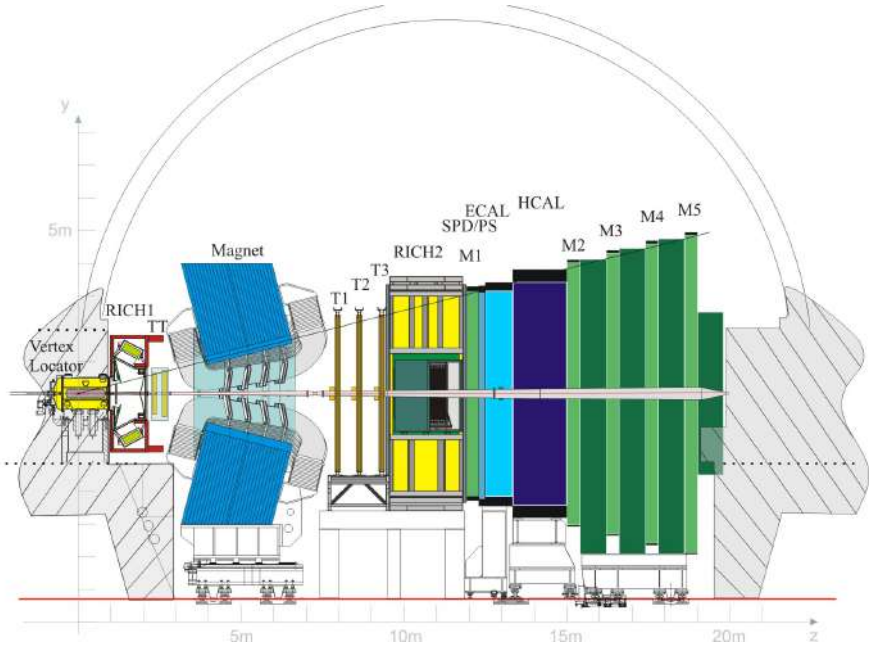


Figure 80: Overview of LHCb detector

FRI 27th June
Today was dedicated to testing updated preselection cuts and analyzing their impact on the data distribution. I ran several checks on the simulation data to evaluate how the cuts improved signal-to-background separation. The results were promising, though there is still room for optimization. In the evening, I treated myself to an Asian buffet at a restaurant near the LHCb experiment site. It was a great way to relax after a busy week, enjoying a variety of dishes and recharging before moving on to the next steps in my analysis.

MON 30th June
This morning, I reviewed the previous week's work and set goals for improving the machine learning model further. I started testing new hyperparameters in XGBoost and documented their effects on classification accuracy. In the afternoon, I participated in a group discussion with other summer students where we shared updates about our projects and exchanged ideas. It was a productive session that gave me fresh perspectives on how to approach some of the challenges in my analysis.

TUE 1st July
Today, I attended my first lecture in the morning and had the chance to meet new friends from different countries. It was a great opportunity to connect with others and share experiences. In the afternoon, I reported my weekly progress, but I realized that I had used irrelevant features when training the machine learning model. This could lead to inaccurate results in the analysis after training, so I plan to adjust my approach moving forward. In the evening, I joined the welcome party for the summer students, where I met many new people. It was a fun and social event, and it gave me a chance to network and build connections with peers from various backgrounds.

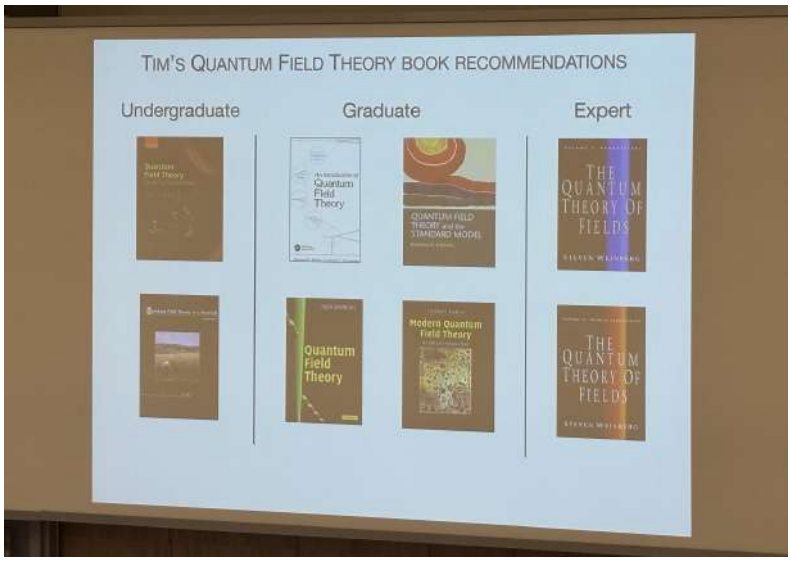


Figure 81: Recommended textbook for Quantum Field Theory from Tim Cohen.

WED 2nd July
Today was another day of continuing my work, focusing on finishing up the pending tasks. The weather was beautiful, as it often is the day after rain. In the evening, after dinner, I took a walk along the hiking path next to CERN and stopped to admire the sunflowers in bloom. It was a peaceful and refreshing way to end the day.

Figure 82: Sun flower field near CERN.

THU 3rd July
Today, I continued working on improving the data analysis, focusing on adjusting the features and model used in the machine learning training. I spent the day reviewing the workflow to ensure it aligns with the updated dataset and identifying areas where further refinements could be made. It was a steady and focused day of work, building on the progress from earlier in the week. In the evening, I cooked dinner for myself—salmon steak and macaroni, a dish I hadn't made in a long time. It felt comforting to enjoy a warm, homemade meal after a full day of work. After dinner, I took a walk to the nearby sunflower field and spent some time admiring the vibrant blooms. It was a simple yet lovely way to end the day, and it felt like one of those small, perfect moments that make a day memorable.

FRI 4th July
Today, I continued working on the same tasks, trying to improve them, but I still encountered issues that caused the machine learning training results to fall short of expectations. Despite the challenges, I remained focused on finding a solution. In the evening, I decided to reward myself for the hard work throughout the week by attending a ceramic painting session. It was a relaxing and creative activity that gave me a break and helped recharge my energy. It also served as a reminder to stay motivated, and I'm ready to tackle the problems again on Monday with renewed determination.

Figure 83: The plate I painted.

MON 7th July
In the afternoon, I met with my supervisor to discuss adjustments to the features and the model used in my machine learning training for the project. We went over the areas that needed improvement, and my supervisor provided some valuable insights on how to refine the process to achieve more accurate results. It was a productive discussion, and I feel more confident about the next steps in my analysis.

TUE 8th July
Today, I got the graph results after making adjustments based on my advisor's suggestions. The results improved, and the peaks are clearer, but they're still not as clear as I'd like. I had a group meeting today to get further advice from my co-advisor. In the evening, I went to pick up the ceramic I had painted at the shop. I realized that I had used too light a color, and the paint wasn't visible at all. However, I know what colors I applied, and anyone who sees my plate can imagine it in whatever color they prefer.

WED 9th July
Today was the final lecture on Theoretical Concepts in Particle Physics. It was another highly engaging and enjoyable session that further fueled my inspiration to pursue studies in theoretical physics. The depth of the topics discussed reminded me why I am so passionate about this field. In the afternoon, I returned to the office to continue working on my project. I attempted to plot the mass distribution of the real data after training the model. Encouragingly, a clear peak corresponding to the mass of the B baryon emerged—it looked better than before, but still not sharp enough to allow for meaningful interpretation. This highlighted the need to explore additional features to improve the model's training and enhance the clarity of the results.

THU 10th July
Today, I switched from analyzing Run 2 data to Run 3. In Run 3, the features and branches differ from those in Run 2, so I spent most of the day matching the branch names in Run 3 to ensure they correspond with the features used for PID, preselection cuts, and training the model in Run 2. In the evening, I decided to cook for myself. I originally planned to make an omelet, but somehow it turned into a classic Thai-style fried egg (kai jeow). It was a fun little cooking experiment and a nice way to end the day.

FRI 11st July
This morning, I visited the Permanent Representative of Thailand in Geneva to discuss and share experiences about my participation in the CERN Summer Student Program, along with other Thai friends who have also joined the program. During the meeting, I also listened to details about the assignments my friends have been working on, each of them contributing to different projects based on their individual interests and expertise. In the evening, I went to the library to spend some quiet time, and later we set out to find a Japanese restaurant for dinner. It took us three tries, as the first two restaurants were either full or closed, but in the end, we finally found a place. The effort was worth it—we enjoyed delicious sushi to wrap up the day.

MON 14th July
Today, I reviewed the Run 2 nTuples and tested tighter BDT and ProbNN cuts to check if the signal peak became clearer. I monitored the impact on statistics, noting that signal loss might require adding more nTuples. I also started checking the Bmass variable (hB_pi0constPVconst_M) after full selection in each of the three CALO regions. This separation should help identify region-dependent effects. Preparing plots for the CALO-specific mass distributions is my next step.

TUE 15th July
I continued with the Run 2 analysis, refining the BDT and ProbNN cut values. I performed preliminary comparisons of CALO region distributions after selection. Observed that some regions have less pronounced peaks, hinting at potential reconstruction differences. Discussed with the team about thresholds for tighter cuts without excessive signal loss. Also prepared a plan to merge more nTuples if needed.

WED 16th July
Today's focus shifted to Run 3 nTuples. I compared MC event numbers across different data blocks to spot inconsistencies. Re-applied ID and preselection to align with previous setups. Investigated differences in feature/branch names, as they might cause issues in re-running the BDT training. Also made notes on potentially selecting a tighter BDT cut for Run 3.

THU 17th July
I started re-training the BDT for Run 3, carefully checking that input variables matched expected formats. Identified that some branches had inconsistent naming between data and MC, requiring mapping adjustments. Began looking into ProbNN variable distributions to decide on optimal cuts. Planned to later compare the Bmass variable across CALO regions and under different BDT/ProbNN thresholds.

FRI 18th July
I completed the first full BDT training for Run 3 with updated feature mappings. I chose an initial tight BDT cut for testing purposes. I began a systematic comparison of Bmass shapes for various ProbNN and BDT thresholds. I compared these results across CALO regions and tested different B mass constraints. I recorded all parameters and planned to cross-check sideband distributions next week.

MON 21st July
I prepared for limited MC statistics by planning to merge blocks 7 and 8 proportionally for both MC and data. I defined the effective photon opening angle variable based on the provided documentation and prepared to add it to the BDT input list. I considered splitting the nTuples into three CALO regions before merging them back. I confirmed that B_OwnPV_MASS would be used for sideband definitions. I also reviewed its equivalence to the Run 2 mass variable. Another activity I did was take a group photo of this year's CERN summer students.

Figure 84: A group photo of ∼300 CERN summer students.

TUE 22nd July
I received new root files produced with an alternative to B_ALLPV_IP. I began adding the effective photon opening angle variable to the ntuples and tested its potential impact in the BDT input. I confirmed that in Run 3, B_OwnPV_MASS should be used as the equivalent mass variable to hB_pi0constPVconst_M. I prepared the steps for implementing new ProbNN cuts. I also checked sideband definitions for consistency.

WED 23rd July
I added the ProbNN variables to the BDT input list and removed the isNotH/isNotE cuts from the BDT. I plotted signal MC and background sideband data in the same figure to examine separation. I tested cut values in the range of 0.4–0.6 to see if they improve discrimination. I observed that moderate cuts could provide a good balance between efficiency and purity. I documented these results for later comparison.

THU 24th July
I focused on tightening the BDT cut to values above 0.9 to test its effect on purity. I checked the selected events in all CALO regions and found that the inner region still had no visible peak. I concluded that reconstruction issues might be affecting the inner region performance. I recorded region-specific notes to guide later optimizations. I prepared histograms for comparison across all regions.

FRI 25th July
I checked the π^0 mass distributions and compared them with the reference PDF values for calibration. I investigated the effect of splitting by the nPVs variable on event quality. I confirmed that the number of primary vertices had a small but noticeable influence on results. I prepared summary plots for discussion with the team. I planned the next week's work based on today's findings. I also went to Lake Geneva to see my senior from Stockholm. We walk along the lake. Fortunately, we saw the rainbow in Jet d'Eau.

Figure 85: Jet d'Eau de Genève

MON 28th July
I began the week by reviewing the results from the previous set of cuts. I tested the merged MC blocks and confirmed they matched the expected proportions relative to the data. I adjusted the binning of the photon opening angle variable to improve its separation power in the BDT. I planned to merge CALO-split datasets for a unified training approach. I documented this preparation for training.

TUE 29th July
I re-trained the BDT with the updated input list, which now includes the photon opening angle and ProbNN variables. I tested sideband definitions using B_OwnPV_MASS limits ($mB_0 < 5050$ and $mB_0 > 5600$). I plotted the signal MC and background sideband distributions to evaluate discrimination after adding the new variables. I observed noticeable improvement in separation. I saved all output plots for reference.

WED 30th July
I validated the performance in the inner, middle, and outer regions after applying the updated cuts. I found that the middle region displayed the most stable signal peak, while the outer region maintained higher statistics but a broader mass distribution. I documented these differences for further optimization. I prepared region-specific efficiency and purity tables. I also archived the training logs for reproducibility.

THU 31st July
I compared π^0 mass distributions after applying final BDT and ProbNN cuts to the PDF reference. I noticed small shifts in peak positions that might require calibration. I re-checked the effect of splitting by nPVs and confirmed a small dependence on the number of vertices. I recorded these results in the analysis notes. I prepared calibration steps to address the observed shifts.

FRI 1st AUG
I finalized the weekly review by compiling dependencies of B mass and π^0 mass on cut values. I created overlay plots comparing variations across CALO regions, BDT/ProbNN thresholds, and B mass constraints. I summarized the findings in a presentation-ready format. I uploaded plots and documentation for the group meeting. I outlined the next steps for selection optimization. It was Swiss National Day. I went to Jet d'Eau to join the celebration.

MON 4th AUG
I presented the progress of my work so far to the team. I focused on fitting the π^0 mass distribution using two components: a linear background and a crystal ball function for the signal. I confirmed that this approach gives a reasonable description of the peak. I noted areas where further refinement may improve the fit quality. I prepared follow-up tasks for the next few days.

TUE 5th AUG
I started working with data and MC from the newest analysis production. I accessed the production link and checked the "Created" and "Updated" columns to separate the new files from the old ones. I planned to use all available data in blocks 7 and 8 for the analysis.
I explored ways to make the root files smaller by selecting only the variables I actually need. I reviewed a shared code snippet to adapt it for this analysis. I also went to the library and borrowed a book to read before bed.

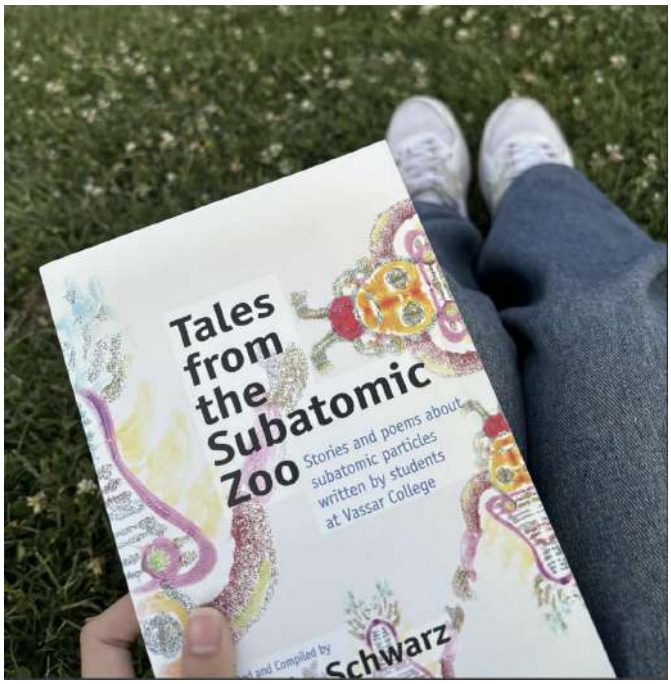


Figure 86: The book I borrowed from CERN library

WED 6th AUG
I joined the student session, which is a project presentation from summer students, in the morning. I began BDT training using only the upper sideband to include only the combinatorial background in the training sample. I replaced the variable B_MINBPVIPCHI2 with B_OWNPVIPCHI2 in the training configuration. I also added the requirement B_OWNPVIPCHI2 < 10 to the preselection cuts. I prepared to check the ProbNN distributions after applying the BDT cut. I noted that these changes should improve background rejection.

THU 7th AUG
Today, I woke up earlier than usual because I had a presentation at the summer student session. My presentation was designed to be a friendly version for an audience with different background knowledge. I gave a brief talk about the interest and challenges of π^0 reconstruction, introduced the concept of BDT (Boosted Decision Tree), explained our objective of discovering new insights from this puzzle, and summarized what I have done so far along with the next steps of my research. In the afternoon, I returned to my office to recheck and continue my work, but there was nothing particularly new to do.

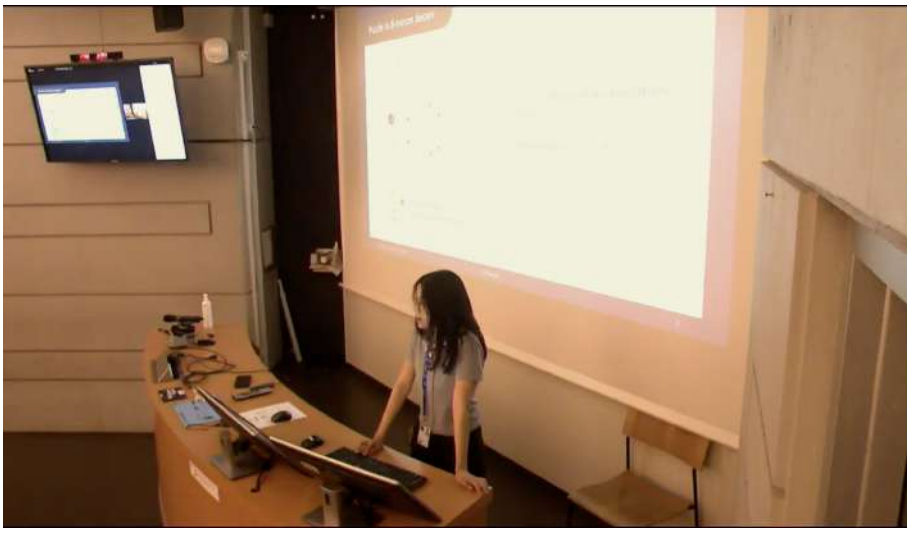


Figure 87: I was giving my presentation in the summer student session.

FRI 8th AUG
I examined the ProbNN distributions after applying the BDT cut to evaluate the separation between signal and background. I produced Dalitz plots for $m(K^+\pi^0)$, $m(\pi^-\pi^0)$, and $m(K^+\pi^-)$ to look for the K^* and ρ resonances. I identified possible structures in the plots that may correspond to these resonances. I planned to make B-mass and π^0 mass plots after these resonance-related changes. I also considered producing separate plots for the outer, middle, and inner regions.

MON 11st AUG
Today I spent more than two hours downloading data from Run 3. However, I accidentally mistyped the preselection code, which made the downloaded data unusable. Therefore, I switched to plotting the mass distribution using Run 2 data instead, since it required less time and I needed results for the group meeting. The AUC and BDT response for run2 turned out very well. After applying a tight BDT cut following the training, I clearly observed the peak of the π^0 mass. When I plotted the B^0 mass, the peak also became clear across all ECAL region. My supervisor then suggested trying to apply the same tight BDT cut for Run 3.

TUE 12nd AUG
In the morning, I spent time downloading the updated Run 3 data, which included our desired feature, $nPVs$, to make the dataset cleaner. I also worked on drafting my presentation for the upcoming LHCb summer student session. In the afternoon, I resumed the Run 3 analysis, and fortunately, the results improved after applying tighter cuts and a low $nPVs$ selection. Although the peak was not as sharp as in Run 2, a clear mass shift could be observed. The AUC and BDT response for this run were good, but the pull of the fit was imbalanced, which may have been caused by issues with the electromagnetic calorimeter calibration.

WED 13rd AUG
Today, I started preparing the slides for my upcoming presentation. I wanted to make the content more technical, focusing on the BDT's role in the analysis and the π^0 reconstruction. I carefully selected plots and results that best represented the progress of my work. Balancing technical depth with clarity took some time, but I managed to draft a solid outline. In the afternoon, I reviewed the structure to ensure that the flow of the talk would be smooth and easy to follow.

THU 14th AUG
I continued working on the slides by adding more detailed explanations of the methodology and results. I included descriptions of the variables used in the BDT and comparisons between Run 2 and Run 3 performance. To make the slides clearer, I adjusted the design and arranged the order of the figures to match the storyline. I also added short captions under the plots so the audience could understand them quickly. By the end of the day, the slides were more complete and professional.

FRI 15th AUG
I sent my draft slides to my supervisor for review. Before sending, I checked everything again to make sure the content was consistent and well-organized. My supervisor provided feedback suggesting that some explanations should be made more concise and certain plots should include more detail. I immediately applied these suggestions by refining the text and highlighting the fits of the π^0 and B^0 mass distributions. I also improved the comparison between Run 2 and Run 3 results. These changes made the slides stronger and more suitable for the presentation.

MON 18th AUG
Today, I received additional advice from two co-advisors that I should reduce the number of slides in my presentation to fit within the allocated time. Some slides were carefully removed, and content was slightly rearranged to make the talk flow more smoothly. This adjustment helped the presentation become more concise and focused. For the rest of the day, I dedicated my time to preparing the script for tomorrow's presentation. I practiced explaining the key points, and I made sure the timing would be appropriate. By the end of the day, I felt more confident about delivering the talk.

TUE 19th AUG

Today, I presented my research at the LHCb Summer Student Session. The presentation was more technical than the previous event, since the audience had a strong background in particle physics. I explained the challenges of π^0 reconstruction, introduced the Boosted Decision Tree (BDT), and described how it helps to distinguish signal from background. I also highlighted the progress I have made so far and outlined the next steps of my research. Although I stumbled and spoke a little hesitantly, the talk was completed successfully. I realized that technical presentations require a balance of detail and clarity. This experience gave me confidence and motivation to improve my presentation skills for future meetings.

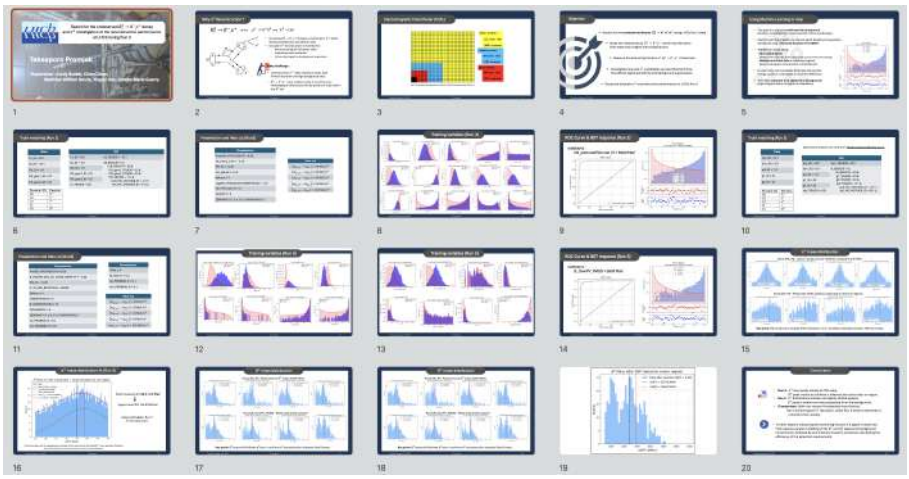


Figure 88: My presentation slide for specific audience in LHCb summer student session.

WED 20th AUG

After finishing my presentation yesterday, today I focused on reviewing my analysis results before including them in the final report. I carefully re-checked the plots of π^0 and B^0 mass distributions under different BDT cuts to confirm their consistency. I also examined the effect of low $nPVs$ selection once more to ensure the improvement was robust. In the afternoon, I reorganized the figures and made notes on which ones would be most suitable to highlight in the report. This preparation gave me a clearer plan for writing the results section.

THU 21st AUG

I continued validating the results, concentrating on comparisons between Run 2 and Run 3. I checked the signal peaks in each dataset and confirmed that the mass shift in Run 3 was consistent with what I observed earlier. I also looked into the ProbNN distributions again to decide which version of the plots should be used in the final draft. In the afternoon, I cross-checked the fitting outputs to ensure that the captions and descriptions would match correctly in the report. The process was time-consuming but important for accuracy.

FRI 22nd AUG

Most of today was spent on final checks of the plots and tables that will go into the report. I reviewed the Dalitz plots and resonance structures once more, making sure the selections applied were consistent across all regions. I also double-checked that the BDT response and AUC values were properly documented. In the afternoon, I created a clean set of figures to be inserted directly into the report next week. By the end of the day, I felt confident that the results were reliable and ready for writing.

MON 25th AUG
Today, I focused on writing my final report. I organized the sections and checked the structure carefully. I made sure the introduction explained the purpose of my work clearly. I also reviewed my notes from previous weeks to include the important results. Although it was a quiet day, I made steady progress.

TUE 26th AUG
I continued writing my report, focusing on the methodology part. I described data samples, selection cuts, and the BDT training steps. I tried to keep the explanations precise but still understandable. The writing took more time than expected, but I am satisfied with the details. This section is almost complete now.

WED 27th AUG
Most of today was spent on writing the results and discussion section. I included the plots I had produced earlier and explained their significance. I also compared the performance between Run 2 and Run 3. The writing process reminded me of the progress I had made over the summer. It was a productive but tiring day.

THU 28th AUG
I continued working on the conclusion and final editing of my report. I summarized the key findings and emphasized possible improvements for future studies. Following my supervisor's suggestion, I added more related plots to strengthen the discussion, especially comparisons between the mass distributions of raw data and those obtained after applying BDT cuts. These additional figures clearly showed how the selection improved the signal visibility and provided stronger support for my analysis. I also checked formatting, references, and figure captions to ensure everything was consistent. By the end of the day, the report was nearly ready for submission.

FRI 29th AUG
Today was my last day of contract. There was not much work to do except submitting my final report to the CDS system, returning the room key, and visiting the secretary's office to complete the contract. I also took the opportunity to say goodbye to my supervisor. It felt a little sad to leave, as this journey has been an important part of my life. At the same time, I am grateful for all the experiences and knowledge I have gained. This program gave me not only academic skills but also valuable life experiences. I will remember these moments and carry them into my future work.

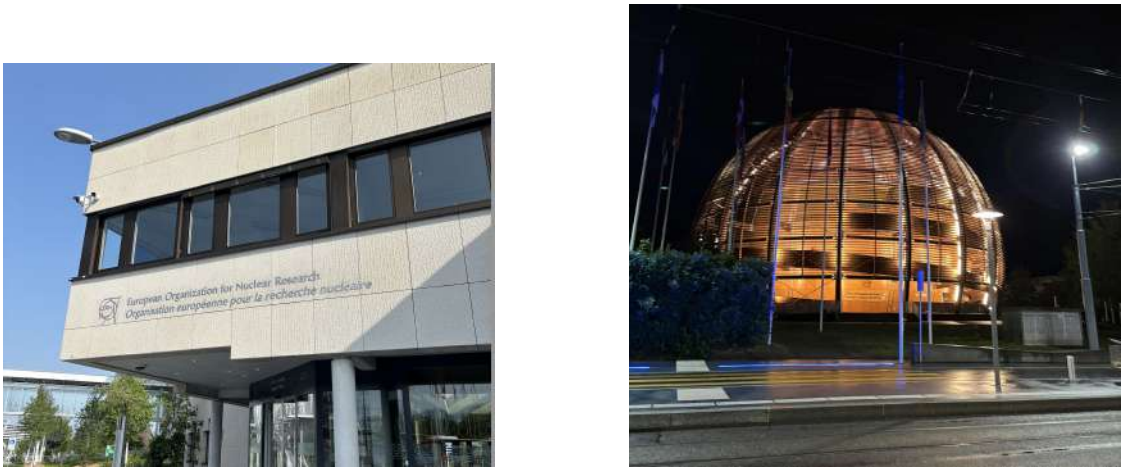


Figure 89: Building 33 (left) and The Globe of Science and Innovation (right)

2 Biography

My name is Taksaporn Promjak. I was born and raised in Chiang Rai, Thailand, where I lived until the age of eighteen. I then moved to Nakhon Ratchasima to pursue my education at the School of Physics, Institute of Science, Suranaree University of Technology.

During my school years, I participated in many science-related activities, although I did not yet have a clear academic goal. This changed when I had the opportunity to attend my first summer school on particle physics, which opened new scientific perspectives and deeply inspired me. That experience sparked my strong interest in particle physics and motivated me to pursue this field further at the university level.

At Suranaree University of Technology, supported by a DPST scholarship, I chose to focus on theoretical particle physics, taking specialized courses and engaging in research in this area. These experiences have strengthened my academic foundation and confirmed my passion for this field.

Looking ahead, I aspire to pursue a Ph.D. in physics and to continue my research career in particle physics, contributing to the advancement of knowledge in this field.

CERN SUMMER STUDENT PROGRAMME 2025

รายงานการเข้าร่วมโครงการนักศึกษาภาคฤดูร้อนเซิร์น